SoK: Metric Differential Privacy in Theory and Practice

Xinpeng Xie
xinpengxie@my.unt.edu
University of North Texas
Department of Computer Science and Engineering

Chenyang Yu
chenyangyu@my.unt.edu
University of North Texas
Department of Computer Science and Engineering

Yan Huang
yan.huang@unt.edu
University of North Texas
Department of Computer Science and Engineering

Yang Cao
cao@c.titech.ac.jp
Institute of Science Tokyo
Department of Computer Science

Chenxi Qiu*
chenxi.qiu@unt.edu
University of North Texas
Department of Computer Science and Engineering

Abstract

Metric Differential Privacy (mDP) extends classical differential privacy (DP) by replacing Hamming adjacency with application-aware distance metrics, which offers utility-preserving protection for structured and continuous data including locations, trajectories, images, and text embeddings.

This Systematization of Knowledge (SoK) paper synthesizes a decade of progress (2013–2025), clarifying mDP's foundations and its connections to central and local DP, and surveying three principal mDP mechanism families: *homogeneous distance mechanisms* (e.g., Laplacian noise), *non-homogeneous distance mechanisms* (e.g., Exponential mechanism), and *optimized perturbation mechanisms*. We organize applications across geo-location privacy, text and embeddings, image and voice protection, graphs and network telemetry, and federated/edge settings. We also surface open challenges, including robust composition and adversarial modeling, context-adaptive privacy, high-dimensional scalability, and principled geometry-aware trade-off bounds, and distill practical guidance for selecting metrics, mechanisms, and metrics of utility. The goal is a unified reference and roadmap for deploying scalable, utility-preserving metric privacy in real-world systems.

Keywords

Metric Differential Privacy, Differential Privacy, Systematization of Knowledge

1 Introduction

Data analysis increasingly relies on sensitive user information, raising a fundamental tension between privacy protection and utility. *Differential Privacy (DP)* [1] provides a principled solution by ensuring that query outputs are statistically stable under small changes to the input dataset, typically via calibrated randomness. In the classical formulation, neighboring datasets differ in one record, commonly formalized through Hamming distance. This binary notion of adjacency is effective for tabular data, but it can be too coarse for structured or continuous domains, where record differences naturally admit magnitudes. By assigning the same privacy cost to all record substitutions, classical DP cannot distinguish small, local perturbations from large semantic changes, which can lead to unnecessary utility loss.

Many real-world applications instead require more nuanced privacy notions that better reflect relationships between data points. For geo-location data, for example, revealing whether a user is within 1 kilometer versus 100 kilometers of a point conveys vastly different levels of sensitive information [3]. In text and image analysis, semantic or perceptual similarity often matters more than a strict binary difference at the record level. Classical DP, which treats all records as equally distinct, is not designed to capture such subtleties. To address this gap, *Metric Differential Privacy (mDP)* [4] generalizes DP by replacing the Hamming distance with application-specific metrics on the input domain. Incorporating domain-aware distance measures increases both the flexibility and the expressiveness of DP, and has enabled its application across a broad range of settings, including geo-location obfuscation [5], word embeddings [6], speech processing [7], and image protection [8, 9].

Since its introduction in 2013, mDP has attracted significant research attention. Figure 1 illustrates the historical growth of publications from 2013 to 2025, with stacked bars that show contributions across domains including location, text, image, and voice. The trend underscores the rising influence and broadening adoption of mDP, as early dominance in location privacy gradually gave way to more diverse applications such as text and image analysis, voice protection, and graph-based privacy. This evolution highlights mDP's expanding impact and its versatility in addressing privacy needs across modern data ecosystems.

Despite substantial progress in the theoretical foundations, algorithmic advances, and practical applications of mDP, to the best of our knowledge there is still no Systematization of Knowledge (SoK) paper dedicated specifically to this topic. Existing SoK/surveys focus primarily on classical DP and local differential privacy (LDP) and mention mDP only briefly. For example, [10] reviews DP- and LDP-based techniques for protecting unstructured data such as images, audio, and text, emphasizing utility challenges and trade-offs but touching on mDP only in passing. Similarly, [11] examines

*Corresponding author.

Proceedings on Privacy Enhancing Technologies 2026(4), 415–437

https://doi.org/10.56553/popets-2026-0128

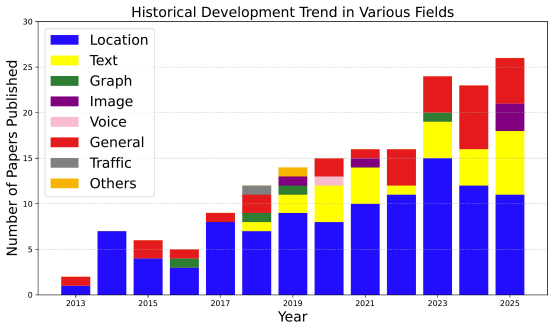


Figure 1: Growing adoption of mDP across multiple application domains.

scenario-specific adaptations of DP for local and statistical data privacy, while again treating mDP as a minor theme. The two existing SoKs on DP are similarly tangential to the goals of this paper. Desfontaines and Péjó [12] systematize the rapidly proliferating family of DP *variants* (e.g., Rényi-DP, zCDP, Gaussian-DP, LDP, group privacy) under a unifying lens of seven dimensions, but treat mDP only as one generalization axis and do not provide a mechanism-level taxonomy. Tschantz *et al.* [13] reframe DP as a *causal* property and clarify what privacy notions different DP variants formalize, which is largely orthogonal to mechanism design. The closest existing comparative work in the location-privacy subdomain is the "Methods for Location Privacy" overview of Chatzikokolakis *et al.* [14], which contrasts mDP-based mechanisms with k-anonymity, dummies, and cloaking; however, it is a domain-specific comparison rather than a systematization of mDP as a general privacy framework. Table 1 summarizes the scope and mDP coverage of these works alongside ours. As a result, these works largely overlook the distinctive theoretical formulations, mechanism designs, and application-driven innovations that characterize mDP. The absence of a unified reference makes it difficult for new researchers to navigate the literature and for practitioners to compare and select appropriate mechanisms. This gap motivates the need for a dedicated, systematic review of developments and challenges unique to mDP.

To address this gap, we present the *first* ***mechanism-level*** *SoK of metric differential privacy* (2013–2025), organized around a unified design taxonomy: *homogeneous distance mechanisms* (HDM), *nonhomogeneous distance mechanisms* (NHDM), and *optimized perturbation mechanisms* (OPM). Our central thesis is that the apparently diverse landscape of mDP designs is best understood as instances of these three families and their hybrids; this lens threads through Sec. 3 (mechanisms), Sec. 4 (applications, summarized in Table 4), and Sec. 5 (open problems). Our contributions are threefold:

- **First mechanism-level taxonomy of mDP.** We propose the HDM/NHDM/OPM partition together with hybrids, and show via a 5-point worked example (Sec. 3) that the three families produce concretely different output distributions on the same physical scenario, isolating the role of *mechanism choice* from the role of *metric choice*, a distinction not made explicit in prior surveys.

Table 1: Comparison of related SoKs/surveys with our work.

Work	Type	Primary scope	Mechanism-level mDP taxonomy
Desfontaines and Péjó [12]	SoK	DP variants	No
Tschantz et al. [13]	SoK	DP as a causal property	No
Chatzikokolakis et al. [14]	Survey	Location privacy	Partial
Zhao and Chen [10]	Survey	DP/LDP for unstructured data	No
Zhao et al. [11]	Survey	Scenario-specific DP/LDP	No
This work	**SoK**	**mDP across domains**	**Yes (HDM/NHDM/OPM)**

Table 2: Core notations used throughout the paper.

Symbol	Meaning
$\mathcal{X}$	Input domain (data universe)
$\mathcal{Y}$	Output (response) space
$d_{\mathcal{X}}$	Metric defined over $\mathcal{X}$
$\mathcal{M} : \mathcal{X} \to \mathcal{Y}$	Randomized mechanism
ε	Privacy budget controlling indistinguishability
δ	Slack parameter (for approximate mDP variants)
$P_{\mathcal{M}(x)}$	Output distribution of $\mathcal{M}$ on input x
$u(x, y)$	Utility or score function (for EM and optimization)
Δf	Sensitivity of query f under $d_{\mathcal{X}}$
η	Random noise variable added by mechanism

- **A cross-domain pattern surfaced by the taxonomy.** Organizing the application literature along HDM/NHDM/OPM (Table 4) reveals a previously-unstated regularity: HDM dominates continuous symmetric domains (geo-coordinates, embeddings); NHDM dominates discrete or structured vocabularies (text, road points, graphs); OPM is forced wherever task utility is direction-dependent, asymmetric, or constraint-driven. This pattern is structural, not historical.
- **Four mDP-specific research frontiers.** We identify and develop four directions where classical ε-DP machinery does not directly transfer: composition under heterogeneous metrics, privacy amplification via shuffling/subsampling without truncation slack, iterative Bayesian update as the correct local-model estimator (rather than matrix inversion), and the gradient-stage gap blocking a true mDP-SGD. All four are downstream of the OPM frontier surfaced by the taxonomy.

The objective of this paper is to *serve as a foundational reference for researchers, practitioners, and system designers who seek to navigate the evolving landscape of metric-based privacy protection and contribute to the next generation of privacy-preserving technologies.*

The remainder of the paper proceeds as follows. Sec. 2 relates mDP to global and local DP; Sec. 3 systematizes HDM, NHDM, and OPM; Sec. 4 surveys applications; Sec. 5 outlines open problems; and Sec. 6 concludes.

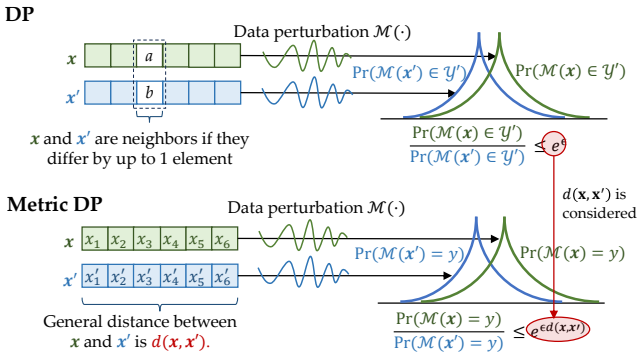


Figure 2: DP vs mDP

2 From DP to mDP

In general, we model a randomized mechanism as a *probabilistic mapping* $\mathcal{M} : \mathcal{X} \to \mathcal{Y}$ from an *input domain* $\mathcal{X}$ to a *perturbed (output) domain* $\mathcal{Y}$. The input space $\mathcal{X}$ is equipped with a distance function $d_{\mathcal{X}} : \mathcal{X}^2 \to \mathbb{R}_{\geq 0}$, and we write $d_{\mathcal{X}}(x, x')$ for the distance between any $x, x' \in \mathcal{X}$. The domain $\mathcal{X}$ and the metric $d_{\mathcal{X}}$ are public; what privacy must protect is the realized *secret input* $x \in \mathcal{X}$ corresponding to a particular user. We use "input domain" to refer to $\mathcal{X}$ as a whole and "secret input" for a specific element $x \in \mathcal{X}$; we avoid the phrase "secret domain" (which conflates the two) for the rest of the paper.

As illustrated in Figure 2, in its original formulation [16], DP requires that the randomized mechanism $\mathcal{M}$ make the outputs $\mathcal{M}(x)$ and $\mathcal{M}(x')$ nearly indistinguishable whenever x and x' are *neighboring databases*, i.e., databases that differ in exactly one entry (Hamming distance one). Formally,

$$\Pr[\mathcal{M}(x) \in \mathcal{Y}'] \leq e^{\varepsilon} \Pr[\mathcal{M}(x') \in \mathcal{Y}'], \quad \forall \mathcal{Y}' \subseteq \mathcal{Y}, \tag{1}$$

where $\varepsilon > 0$ is the *privacy budget*, and x and x' are *neighboring* inputs under a specified adjacency relation.

For self-containment, we restate the two pre-existing notions whose generalization mDP captures, before introducing mDP itself.

Definition 2.1 (Central ε-DP [16]). *A randomized mechanism $\mathcal{M} : \mathcal{D} \to \mathcal{Y}$ operating on databases $D \in \mathcal{D}$ satisfies ε-differential privacy if, for every pair of* neighboring *databases $D, D' \in \mathcal{D}$ (those differing in exactly one record) and every measurable output set $\mathcal{Y}' \subseteq \mathcal{Y}$,*

$$\Pr[\mathcal{M}(D) \in \mathcal{Y}'] \leq e^{\varepsilon} \Pr[\mathcal{M}(D') \in \mathcal{Y}'].$$

Intuition. Central DP assumes a *trusted curator* that runs $\mathcal{M}$ on the full database; the same multiplicative bound e^{ε} applies to every neighboring pair regardless of how different they are in any other sense.

Definition 2.2 (ε-LDP [22]). *A randomized mechanism $\mathcal{M} : \mathcal{X} \to \mathcal{Y}$, applied independently by each user to her own datum $x \in \mathcal{X}$, satisfies ε-local differential privacy if, for every pair of records $x, x' \in \mathcal{X}$ and every measurable output set $\mathcal{Y}' \subseteq \mathcal{Y}$,*

$$\Pr[\mathcal{M}(x) \in \mathcal{Y}'] \leq e^{\varepsilon} \Pr[\mathcal{M}(x') \in \mathcal{Y}'].$$

Intuition. LDP assumes *no trusted curator*: each user perturbs her datum locally, and the same e^{ε} bound must hold for *every* pair of records — making LDP typically far noisier than central DP at matched ε (Fig. 2 previews how mDP relaxes this uniform bound to a metric-graded one).

mDP — introduced by Chatzikokolakis et al. [4] under the original name *d-privacy* (or *$d_{\mathcal{X}}$-privacy* when emphasizing the metric on the input domain [20, 83])[1] generalizes the DP constraint by enforcing indistinguishability according to an arbitrary metric $d_{\mathcal{X}}$ defined on the input domain $\mathcal{X}$. Rather than requiring a fixed bound only for adjacent inputs, the likelihood-ratio bound scales with the distance between secrets:

$$\Pr[\mathcal{M}(x) \in \mathcal{Y}'] \leq e^{\varepsilon d_{\mathcal{X}}(x,x')} \Pr[\mathcal{M}(x') \in \mathcal{Y}'], \tag{2}$$

$\forall x, x' \in \mathcal{X}, \ \forall \mathcal{Y}' \subseteq \mathcal{Y}$. We formalize mDP as follows.

Definition 2.3. $(\varepsilon, d_{\mathcal{X}})$-*mDP.* *A randomized mechanism $\mathcal{M} : \mathcal{X} \to \mathcal{Y}$ satisfies $(\varepsilon, d_{\mathcal{X}})$-mDP if, for all $x, x' \in \mathcal{X}$ and all $\mathcal{Y}' \subseteq \mathcal{Y}$, the constraint in Eq. (2) holds.*

Intuitively, mDP requires that *nearby secrets* (under $d_{\mathcal{X}}$) induce output distributions that are correspondingly close: the multiplicative gap is bounded by $e^{\varepsilon d_{\mathcal{X}}(x,x')}$. This is particularly useful for continuous or structured domains, where differences between secrets naturally admit magnitudes, and treating all changes as equally sensitive can be overly restrictive.

Classical ε-DP in Eq. (1) is recovered as a special case by letting the adjacency relation be induced by the Hamming distance d_{h}, i.e., declaring x and x' to be neighbors when $d_{\mathrm{h}}(x, x') = 1$; in this case, Eq. (2) reduces to the fixed bound in Eq. (1) for any single-record change.

mDP vs. LDP and central DP. Before turning to mechanisms, we clarify how mDP relates to the two notions of Definitions 2.1–2.2. Our position is that *mDP is conceptually orthogonal to the local/central distinction, but practically most closely associated with the local model for both historical and utility reasons.*

Conceptually, mDP only specifies how indistinguishability scales with a metric on the input domain; it does not commit to where the noise is added. Both local and central instantiations exist in the literature. As a special case, taking $d_{\mathcal{X}}$ to be the Hamming distance and applying mDP record-wise recovers central ε-DP (Definition 2.1); applying mDP at the level of an individual record, on the user's device, with the same fixed bound on every pair, recovers ε-LDP (Definition 2.2).

Practically, mDP was first motivated as a relaxation of LDP to address its utility collapse on individual data: Andrés et al. [5] observed that applying standard DP to a single user's location "makes it impossible to communicate any useful information to the service provider," and introduced Geo-Indistinguishability as a metric-graded relaxation. The bulk of mDP work that followed —

[1] The umbrella term "metric differential privacy" (mDP) emerged in subsequent literature as a more accessible name; the original 2013 formulation [4] used "*d*-privacy", and the term "mDP" is now standard in surveys and titles (see, e.g., [3, 58]). The closely-related notion of *Lipschitz privacy* [19] is *strictly stronger* than mDP: it requires the log-density of the mechanism's output to be Lipschitz in the input (equivalently, a pointwise gradient bound where the density is differentiable). Mechanisms with absolutely continuous, smooth output densities (e.g., Laplace, Gaussian) satisfy both Lipschitz privacy and mDP at the same ε, but mechanisms with discontinuous densities — notably the staircase mechanism, which still satisfies ε-DP — satisfy DP/mDP yet fail Lipschitz privacy [19]. We therefore treat them as related-but-distinct notions throughout this paper, and our use of "mDP" refers exclusively to the measurable-set definition in Eq. (2) below.

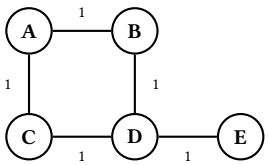


Figure 3: Running example used throughout Sec. 3: a 5-point toy road network with all edge lengths = 1 km. Cartesian coordinates: $A = (0, 1)$, $B = (1, 1)$, $C = (0, 0)$, $D = (1, 0)$, $E = (2, 0)$. Pairwise distances under three metrics are listed in Table 3.

geo-indistinguishability for location [5, 110], text and embedding sanitization [20, 28], trajectory release [23, 44] — consequently sits in the local model.

Central instantiations of mDP do exist. Kamalaruban et al. [83] explicitly study $d_{\mathcal{X}}$-private mechanisms for linear queries over histograms in the centralized model, tailoring Laplace and related mechanisms to a metric on the data domain so that more sensitive attributes receive stronger protection. Imola et al. [58] develop user-level mDP under the Earth Mover's Distance and provide mechanisms applicable to both central and local deployments, making the orthogonality explicit. mDP also retains post-processing immunity and sequential composition like classical DP [3]; we treat composition formally in Sec. 5.3.

Compared to LDP, mDP's metric-graded noise yields three concrete advantages: *(1) better utility on high-dimensional or continuous data*, since LDP's uniform noise scales poorly with dimension while mDP only enforces indistinguishability among nearby points [25, 42]; *(2) alignment with domain-specific similarity* — Euclidean or haversine distance for geo data [5], Word Mover's Distance for text [28], perceptual or feature-space metrics for images [9] — so privacy budget tracks meaningful structure rather than data-agnostic randomization; and *(3) finer privacy–utility control*, by tuning $\varepsilon \cdot d_{\mathcal{X}}$ rather than a uniform ε [3]. Table 6 surveys the distance metrics commonly adopted in mDP work, ranging from classical geometric measures (Euclidean, Manhattan, shortest path) through semantic ones (Word Mover's Distance, Wasserstein) to specialized notions for structural and temporal data (node-distance for graphs, angular distance for embeddings and voice, Fréchet for trajectories, temporal distance for timestamped sequences).

Running example. We instantiate each mechanism on the 5-point toy road network of Fig. 3 ($\mathcal{X} = \{A, B, C, D, E\}$ on a 1-km grid). The graph induces three distances (Table 3): Hamming d_h (used by ε-DP); Euclidean $\|\cdot\|_2$ (planar Laplace), e.g., $\|A-D\|_2 = \sqrt{2}$; shortest-path d_g on the road graph (graph-aware EM), giving $d_g(A, D)$=2 because vehicles cannot cut diagonally. The disagreement between $\|\cdot\|_2$ and d_g on $\{A, D\}, \{A, E\}, \{B, C\}, \{B, E\}, \{C, E\}$ is what makes graph metrics meaningful for road networks. We fix $\varepsilon = \ln 2 \approx 0.693$ throughout for arithmetic transparency, conditioning on x=A (or x=D) when computing example output distributions.

3 Key Data Perturbation Mechanisms in mDP

In this section, we review the core mechanisms that have shaped the development of mDP. Particularly, we classify existing perturbation

Table 3: Pairwise distances on the running example (Fig. 3). d_h: Hamming (used by ε-DP); $\|\cdot\|_2$: Euclidean (planar Laplace); d_g: shortest path on the road graph (graph-aware EM). Bold entries highlight pairs where $\|\cdot\|_2$ and d_g disagree.

Pair	A, B	A, C	A, D	A, E	B, C	B, D	B, E	C, D	C, E	D, E
d_h	1	1	1	1	1	1	1	1	1	1
$\|\cdot\|_2$	1	1	$\sqrt{2}$	$\sqrt{5}$	$\sqrt{2}$	1	$\sqrt{2}$	1	2	1
d_g	1	1	**2**	**3**	**2**	1	**2**	1	**3**	1

mechanisms for mDP according to how they use the underlying metric and how the perturbation kernel is specified. Our taxonomy classifies by *design principle*: a non-homogeneous distance mechanism (NHDM) encodes the privacy loss directly as a closed-form, distance-weighted distribution, whereas an optimized perturbation mechanism (OPM) treats the mechanism as the *solution* of an optimization problem with mDP as a constraint. The distinction is conceptually clean but becomes a continuum in practice — e.g., the LP-optimized exponential mechanism of [3] uses a closed-form EM kernel (NHDM) whose parameters are set by an LP (OPM). We treat such cases as *hybrid mechanisms* and discuss them explicitly in Sec. 3.4; works with hybrid designs are tagged by multiple family checkmarks (and a $^{\mathrm{H}}$ marker) in Table 10.

(1) Homogeneous distance mechanisms (Section 3.1). These mechanisms perturb a secret value x in such a way that the resulting output distribution depends solely on the distance between x and the reported output y, and not on their absolute locations in the domain. Formally, the conditional distribution has the form

$$\Pr\big[\mathcal{M}(x) = y\big] \propto \phi\big(d_{\mathcal{X}}(x, y)\big), \tag{3}$$

for some monotone decay function ϕ, making the mechanism fully determined by the metric $d_{\mathcal{X}}$ and the privacy parameter.

(2) Non-Homogeneous Distance Mechanisms (Section 3.2). These mechanisms remain distance-based but are *locally geometry-aware*. When the output domain $\mathcal{Y}$ is finite, a typical instantiation has the form

$$\Pr[M(x) = y] = \frac{\phi\big(d(x, y)\big)}{\sum_{y' \in \mathcal{Y}} \phi\big(d(x, y')\big)}, \tag{4}$$

where ϕ is again a monotone decay function of the distance. Here the numerator depends on $d(x, y)$, while the normalization factor depends on the entire set $\{d(x, y') : y' \in \mathcal{Y}\}$. Consequently, the perturbation distribution reflects the *relative configuration* of the surrounding candidates y' around x, and thus implicitly adapts to local density or neighborhood structure in the input domain.

(3) Optimized Perturbation Mechanisms (Section 3.3). These mechanisms are *distance-constrained but utility-agnostic*. The perturbation probability $z_{x,y} = \Pr[M(x) = y]$ is obtained as the solution of an optimization problem that enforces the mDP constraints

$$z_{x,y} \le e^{\varepsilon d(x,x')} z_{x',y} \quad \text{for all } x, x' \in \mathcal{X}, y \in \mathcal{Y}, \tag{5}$$

while minimizing a general loss (or utility loss) that need not be proportional to $d(x, y)$. In this family, the metric d appears structurally in the privacy constraints, whereas the objective can encode arbitrary task- or semantics-specific preferences over outputs

(LP/QP-based perturbation-matrix constructions are typical examples).

A comparative analysis of these mechanisms is presented in **Section 3.4**. The historical evolution of these three categories is illustrated in Figure 4.

3.1 Homogeneous Distance Mechanisms

The Laplace mechanism, originally formulated for ε-DP via $\mathcal{M}(f(x)) = f(x) + w$ with $w \sim \mathrm{Lap}(\Delta f/\varepsilon)$ and global sensitivity Δf, generalizes to mDP by scaling noise to the metric distance $d_{\mathcal{X}}(x, y)$ instead of Δf [4, 43], giving the homogeneous form $\Pr[\mathcal{M}(x)=y] = \lambda\, e^{-\varepsilon\, d_{\mathcal{X}}(x,y)}$. The most prominent instance is planar Laplace for 2D location data [5], the canonical HDM anchor below.

Anchor: Planar Laplace (Andrés et al., CCS 2013).

Setting. Input domain $\mathcal{X} \subseteq \mathbb{R}^2$ with Euclidean metric $d_{\mathcal{X}} = \|\cdot\|_2$.

Privacy property (geo-indistinguishability). $\mathcal{M} : \mathcal{X} \to \mathcal{X}$ is ε-GeoInd iff $\Pr[\mathcal{M}(x) \in S] \le e^{\varepsilon\|x-x'\|_2} \Pr[\mathcal{M}(x') \in S]$ for all $x, x' \in \mathcal{X}$ and measurable S.

Mechanism (continuous PDF). $\Pr[\mathcal{M}(x)=y] = \frac{\varepsilon^2}{2\pi} e^{-\varepsilon\|x-y\|_2}$, sampled in polar form: draw $\theta \sim \mathrm{Unif}[0, 2\pi)$, $u \sim \mathrm{Unif}[0, 1)$; set $r = -\frac{1}{\varepsilon}\left(W_{-1}(\frac{u-1}{e}) + 1\right)$ via the Lambert-W_{-1} branch; return $y = x + (r\cos\theta, r\sin\theta)$.

Worked example on Fig. 3. At $\varepsilon = \ln 2$, $x = A$: un-normalized weights $\propto e^{-\varepsilon\|A-\cdot\|_2}$ are A:1, B=C:0.500, D:$2^{-\sqrt{2}} \approx 0.375$, E:$2^{-\sqrt{5}} \approx 0.212$, normalizing to A:0.387, B=C:0.193, D:0.145, E:0.082 on the discrete grid. B, C tie because both lie at Euclidean distance 1 from A; the disagreement with d_g surfaces at D, E in the EM anchor below.

Privacy/utility. Planar Laplace is $(\varepsilon, \|\cdot\|_2)$-mDP exactly [5] and is the *optimal* HDM for ℓ_2 identity queries [19, 72]. Two intrinsic weaknesses motivate the next two families: *oblivious-to-discreteness* (samples may land off-manifold), motivating remapping [74] and NHDM; and *direction-agnosticism*, motivating OPM (Sec. 3.3). *Word Mover Laplace* [28] adapts the same recipe to text by replacing $\|\cdot\|_2$ with Word Mover's Distance over embeddings, illustrating that HDM is a recipe parameterized by a metric rather than a single mechanism.

The deeper observation to extract from the literature is that HDM is not a single mechanism but a *recipe*: $\Pr[y|x] \propto e^{-\varepsilon\, d(x,y)}$, optionally post-processed and composed over time. Work beyond the planar-Laplace anchor populates five orthogonal degrees of freedom of this recipe (Table 7). *(A) The metric d itself*, swapped for 3D Euclidean indoors, Word Mover's Distance over text [28], GAN-latent perceptual metrics [9], graph node-distance, or time-series similarity — highlighting that "HDM" is parameterized by a metric rather than a single mechanism. *(B) The privacy budget ε*, made context-dependent through the Eikonal PDE [77] (a first-order PDE that constrains how a spatially-varying $\varepsilon(x)$ must change smoothly to remain consistent with $d_{\mathcal{X}}$), tile-based partitions, semantic land-use levels, per-attribute sensitivities [83], or policy graphs (Blowfish, PGLP) that explicitly enumerate which input pairs must remain indistinguishable. *(C) Sequential composition*, addressed by reusing temporally correlated noise on predictable mobility traces (the foundational predictive mechanism of [23], PETS 2014) and refined by trajectory clustering, adaptive update frequency, and COVID-era contact-tracing deployments. *(D) Theory and post-processing*, including Laplace's MSE-optimality under Lipschitz privacy [19] (later extended to continuous queries [72]), Bayesian remapping with prior [74], and gradual release. *(E) System integration*, where HDM serves as the privacy substrate for mobile-crowdsourcing task assignment, federated-learning group privacy via d-privacy [91], and blockchain-coordinated indoor LBS. The two intrinsic weaknesses of the anchor persist across all five axes: HDM samples may fall *outside* the discrete or structured set the application requires (motivating NHDM, Sec. 3.2) and remain *direction-agnostic* regardless of how task utility shapes the loss (motivating OPM, Sec. 3.3).

Most recently, [46] extended the Laplace-based family by combining mDP with the shuffle model, proposing three constructions — RR-Shuffle, Geo-Shuffle, and SGDL-Shuffle — that share a common design pattern: encode each user's locally-perturbed value as a unary bit vector before shuffling, so that the shuffled output is privacy-equivalent to revealing only the sum of bits. The privacy boost is governed by the previously-unstudied *Symmetric Generalized Discrete Laplace (SGDL)* distribution, which captures how summed unary noise concentrates with the number of users n. This represents the first systematic exploration of shuffle-based metric privacy within the homogeneous distance paradigm; we discuss the mDP-specific obstacles, the largely-unexplored case of subsampling amplification, and the privacy/utility/trust trade-offs surfaced by these mechanisms in Sec. 5.4.

Takeaway. HDM are the simplest and most scalable mDP family: the Laplace closed form depends only on $d_{\mathcal{X}}(x, y)$, sampling is $O(1)$ per release, and they are provably *optimal* for ℓ_2 identity queries [19, 72]. They are the right choice when (i) the input/output domain is continuous, (ii) utility loss is approximately radial in $d_{\mathcal{X}}$, and (iii) noise can be added directly to the data *value* rather than to a selection from a discrete set. They fail when the output must be a valid element of a discrete structured set (a word, a road-network point, a graph state)—motivating NHDM (Sec. 3.2)—or when utility is direction-dependent in ways orthogonal to $d_{\mathcal{X}}$—motivating OPM (Sec. 3.3).

3.2 Non-Homogeneous Distance Mechanisms

Non-homogeneous distance mechanisms extend homogeneous distance mechanisms by making the perturbation distribution depend not only on the distance between the input x and a candidate output y, but also on the *configuration of all candidates around x*. In contrast to homogeneous mechanisms such as Laplace noise, which treat the metric space as locally uniform and assign probabilities based solely on $d_{\mathcal{X}}(x, y)$, non-homogeneous distance mechanisms normalize over a discrete candidate set and thereby adapt to local density and boundary effects.

Anchor: The Exponential Mechanism (McSherry & Talwar, FOCS 2007).

Setting. A finite output domain $\mathcal{Y}$ together with a *utility (score) function* $u : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$, where larger $u(x, y)$ means more desirable. EM picks $y \in \mathcal{Y}$ with probability $\propto \exp(\varepsilon\, u(x, y)/(2\Delta u))$, where $\Delta u = \max_{y, x\sim x'} |u(x, y) - u(x', y)|$ is the score sensitivity [43].

mDP specialization. Set $u(x, y) = -d_{\mathcal{X}}(x, y)$. By the triangle inequality, $|d_{\mathcal{X}}(x, y) - d_{\mathcal{X}}(x', y)| \le d_{\mathcal{X}}(x, x')$, so the standard EM

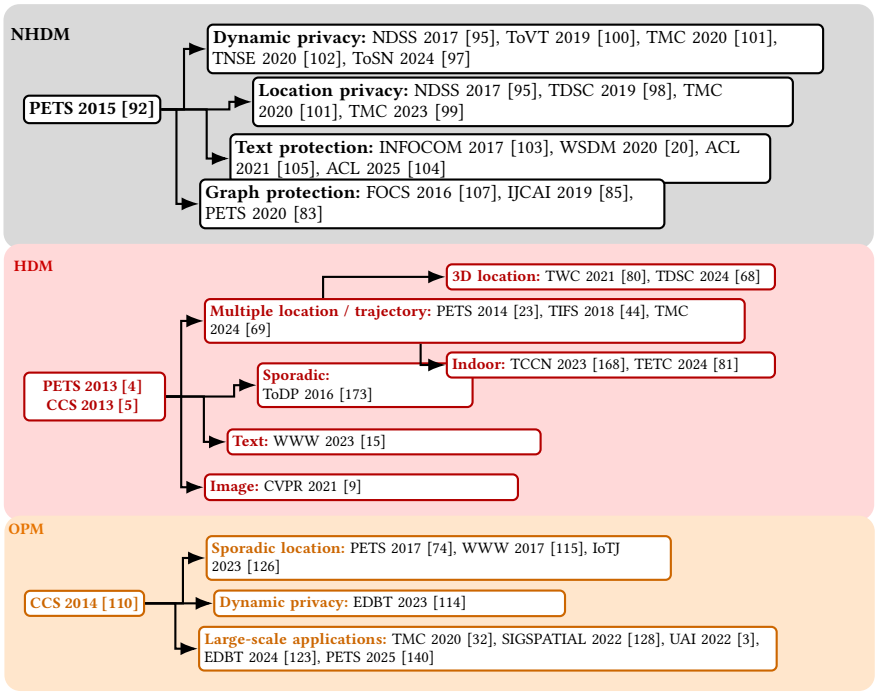


Figure 4: The evolution of three categories of mDP mechanisms: *Homogeneous Distance Mechanisms (HDM), Non-Homogeneous Distance Mechanisms (NHDM),* and *Optimized Perturbation Mechanisms (OPM),* from 2013 to 2025.

proof carries through with sensitivity replaced by $d_{\mathcal{X}}(x, x')$, giving the metric ratio $e^{\varepsilon\, d_{\mathcal{X}}(x,x')}$. The mechanism takes the form

$$\Pr[\mathcal{M}(x) = y] = a_x \exp\!\big(-\tfrac{1}{2}\varepsilon\, d_{\mathcal{X}}(x, y)\big), \qquad a_x = \left(\sum_{y' \in \mathcal{Y}} e^{-\frac{1}{2}\varepsilon d_{\mathcal{X}}(x,y')}\right)^{-1}. \tag{6}$$

Worked example on Fig. 3. At $\varepsilon = \ln 2$, $x = A$, $d = d_g$: unnormalized scores $\exp(-\frac{1}{2}\varepsilon\, d_g(A, \cdot))$ normalize to A:0.306, B=C:0.216, D:0.153, E:0.108. The pedagogically key effect: under planar Laplace's Euclidean metric, $\Pr[D]/\Pr[E] \approx 1.77$ (driven by $\sqrt{5} - \sqrt{2} \approx 0.83$); under EM's graph metric, $\Pr[D]/\Pr[E] \approx 1.41$ (driven by d_g gap $3 - 2 = 1$). A vehicle confined to road segments faces a different privacy–utility trade-off than in straight-line space — the motivation for graph-aware Geo-Graph-Indistinguishability [93].

Elastic-metric variant (NHDM canonical instance). [92] retains the EM kernel form but replaces $d_{\mathcal{X}}$ with an *elastic distinguishability metric* locally rescaled by an estimate of secret-density: in dense regions the metric stretches (stronger privacy), sparse regions are compressed. PRIVIC [50] subsequently re-derived this elastic mechanism as the fixed point of a Blahut–Arimoto iteration (a classic alternating-projection algorithm from information theory, originally designed to compute channel capacity) under an mDP constraint, providing an information-theoretic justification.

Privacy/utility. EM achieves $(\varepsilon, d_{\mathcal{X}})$-mDP by construction, with two strengths over HDM: outputs are guaranteed valid elements of $\mathcal{Y}$ (no off-manifold samples), and the per-input normalizer makes EM *locally density-aware*. Its weakness mirrors HDM's: probability mass is still tied to $d_{\mathcal{X}}(x, y)$ alone, motivating OPM (Sec. 3.3).

A canonical instance is the *Exponential Mechanism (EM)* [43], adapted to mDP via the elastic-metric construction of [92] and information-theoretically reformulated by PRIVIC [50]. EM normalizes over a discrete candidate set, concentrating mass according to a score function while respecting metric-based privacy. Formally, given an input $x \in \mathcal{X}$ and a finite output domain $\mathcal{Y}$, the mechanism selects $y \in \mathcal{Y}$ with probability

$$\Pr\big[\mathcal{M}(x) = y\big] = a_x \exp\!\big(-\tfrac{1}{2}\varepsilon\, d(x, y)\big) \tag{7}$$

The work beyond the EM anchor populates five orthogonal degrees of freedom of the EM recipe, which we organize and detail in Table 8. *(A′) The candidate set $\mathcal{Y}$ and its metric*: shifting from a generic grid or vocabulary to road-graph nodes with shortest-path distance [93, 143], sensitivity-bipartitioned vocabularies [105], hierarchical cluster-then-token sets [104], ordinal subsets [103], anisotropic ellipse-shaped output regions for directional traffic [37], or full word-embedding spaces [20]. *(B′) The score function $u(x, y)$ and budget allocation across structure*: distribution-preserving scores [99]; density-modulated scores [106]; order-preserving scores for cryptographic primitives [108]; linguistic-importance per-token budget weighting (*Spend Your Budget Wisely*) [17]; and PIVE's coupling of EM with an expected-inference-error (EIE, a posterior-based criterion that bounds an adversary's expected Bayesian reconstruction error) constraint as a second criterion [95, 96]. *(C′) Composition over time and dynamic regimes*: regionalized DPIVE that varies budget by spatial cluster [97]; the Eclipse defense against long-term observation attacks [101]; and dynamic-spectrum-sharing variants for evolving auctions [100]. *(D′) Output validity and density-aware adaptation*: cooperative cloaking regions for vehicles [94]; the Truncated EM (TEM) that excludes low-density tail outputs [142]; and

Perturbation-Hidden, which guarantees only *plausible* outputs (no vehicles in lakes) [102]. *(E′) Theory, optimization integration, and system deployment*: the canonical LP+EM hybrid of [3], multi-objective evolutionary EM (MOEA) [109], Lipschitz-extension generalizations to graph statistics [107], EM for graph-infrastructure obfuscation [85], and condensed local DP for small-population datasets [98]. The persistent failure mode across all five axes is shared with the anchor: EM's mass is still tied to $-d(x, y)$ alone, so two outputs at equal distance receive equal probability regardless of how task utility shapes the loss — exactly the limitation that motivates OPM in Sec. 3.3.

NLP worked example: how the same sentence is sanitized under three EM-based text mechanisms. To make the text-domain trade-offs concrete, consider sanitizing the toy sentence "*I love this product*" over a small word-embedding vocabulary $\mathcal{V}$ with semantic distance d (e.g., normalized ℓ_2 on WORD2VEC). All three mechanisms below are EM-NHDM at heart; they differ only in *which candidate set $\mathcal{Y}$ each token sees*.

SanText [105]. Per token w, sample $\tilde{w} \in \mathcal{V}$ from EM with score $-d(w, \cdot)$ over the *entire* vocabulary. Because $\mathcal{V}$ is huge ($\sim 10^5$), the high-density region around *love* contains hundreds of candidates and EM's normalizer disperses probability widely. A typical sample for *love* is a near-synonym (*like, enjoy*) but unrelated frequent tokens are reachable too. Sanitized output: "*I like this item.*"

CusText (cited via the CluSanT paper [104]). Per token w, restrict EM's candidate set to a small *customized* sub-vocabulary $\mathcal{V}_w \subset \mathcal{V}$ (typically the top-k semantic neighbors of w). For *love*, $\mathcal{V}_{love}$ = {*love, adore, like, enjoy, cherish*}, so the replacement is almost surely sentiment-preserving. Sanitized output: "*I adore this product.*"

CluSanT [104]. Two-stage EM. *Stage 1:* pick a semantic *cluster* for w (here, the cluster {*love, adore, like, enjoy*}) using EM under a cluster-level metric. *Stage 2:* pick a token within that cluster using EM under the within-cluster metric. The two stages share the budget $\varepsilon = \varepsilon_1 + \varepsilon_2$. Sanitized output: "*I enjoy this product.*"

Trade-offs. SanText offers full-vocabulary indistinguishability (the strongest privacy guarantee) at the price of frequent semantic drift. CusText is the most utility-friendly but its custom-vocabulary construction is itself a function of the input, raising subtle leakage questions if $\mathcal{V}_w$ is observed. CluSanT interpolates by structuring the candidate space hierarchically, paying a budget split for the two-stage selection. The same observation as in the geo-example holds here: HDM-based mechanisms (Word Mover Laplace) sample noisy *vectors* that may not be valid words at all, while these EM-NHDM mechanisms sample valid tokens directly from $\mathcal{V}$ — which is precisely why the text domain has gravitated toward NHDM rather than HDM.

Takeaway. NHDM (canonical example: EM under $u = -d_{\mathcal{X}}$) replace HDM's "add noise to a value" with "sample from a discrete candidate set", normalizing locally over candidates. This buys two things HDM cannot: (a) outputs are guaranteed valid elements of $\mathcal{Y}$ (no off-manifold points or invalid words), avoiding the lake/word-not-in-vocabulary failure modes of Sec. 3.1; (b) the per-input normalizer adapts to local candidate density. Probability mass nonetheless remains tied to raw distance, however — two outputs at equal $d_{\mathcal{X}}(x, y)$ receive equal probability regardless of task value, which is exactly what motivates OPM (Sec. 3.3). Recent text mechanisms (SanText, CusText, CluSanT) and graph-aware location mechanisms (Geo-Graph-Ind) are all NHDM variants distinguished primarily by which candidate set $\mathcal{Y}$ they feed to EM.

3.3 Optimized Perturbation Mechanisms

While homogeneous distance mechanisms and non-homogeneous distance mechanisms (e.g., Laplace and EM) provide generic constructions for mDP, they tie selection probabilities directly to the perturbation magnitude $d(x, y)$. As a result, they cannot fully capture task-specific or anisotropic utility losses, where perturbations of the same metric distance in different directions may have very different effects. This limitation has motivated a growing line of work on *optimized perturbation mechanisms*, which treat the perturbation kernel as the solution of an explicit optimization problem under mDP constraints.

Given the computational challenges of optimizing a mechanism $\mathcal{M}$ when the input domain $\mathcal{X}$ and output domain $\mathcal{Y}$ are continuous, a common strategy is to discretize both $\mathcal{X}$ and $\mathcal{Y}$ into finite sets [3]. In this setting, the random mechanism $\mathcal{M}$ is represented by a stochastic *perturbation matrix* $\mathbf{Z} = [z_{x,y}]_{(x,y)\in\mathcal{X}\times\mathcal{Y}}$, where each entry $z_{x,y} = \Pr[\mathcal{M}(x) = y]$ denotes the probability of reporting $y \in \mathcal{Y}$ when the true value is $x \in \mathcal{X}$. A standard formulation optimizes $\mathbf{Z}$ by solving the following linear program:

$$\min_{\mathbf{Z}} \quad \sum_{x\in\mathcal{X}} \sum_{y\in\mathcal{Y}} \pi_x \, z_{x,y} \, c_{x,y} \tag{8}$$

$$\text{s.t.} \quad z_{x,y} \le e^{\varepsilon d_{\mathcal{X}}(x,x')} \, z_{x',y}, \quad \forall x, x' \in \mathcal{X}, \ \forall y \in \mathcal{Y}, \tag{9}$$

$$\sum_{y\in\mathcal{Y}} z_{x,y} = 1, \quad \forall x \in \mathcal{X}, \tag{10}$$

$$z_{x,y} \ge 0, \quad \forall (x, y) \in \mathcal{X} \times \mathcal{Y}. \tag{11}$$

Here, $c_{x,y}$ encodes the utility loss incurred when reporting y instead of the true value x, and π_x is a prior distribution over $\mathcal{X}$. The metric $d_{\mathcal{X}}$ appears explicitly in the mDP constraints, while the loss matrix $c_{x,y}$ can express arbitrary, potentially direction-dependent utility, decoupling the notion of utility from pure metric distance.

Anchor: LP-Optimal Geo-Indistinguishable Mechanism (Bordenabe et al., CCS 2014).

Setting. Discrete domains $\mathcal{X}, \mathcal{Y}$ (often $\mathcal{X} = \mathcal{Y}$), a metric $d_{\mathcal{X}}$, a prior π over $\mathcal{X}$, and a *task-specific* utility-loss function $c : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}_{\ge 0}$ that need *not* be a function of $d_{\mathcal{X}}$ alone (e.g., c may be direction-dependent or driven by downstream task error).

LP formulation. The perturbation matrix $\mathbf{Z} = [z_{x,y}]$ is the decision variable; the mDP constraints in Eq. (8) are exactly the privacy bound, and the simplex constraints make each row a probability distribution. The LP has $|\mathcal{X}||\mathcal{Y}|$ variables and $\Theta(|\mathcal{X}|^2|\mathcal{Y}|)$ privacy constraints [110], making direct deployment feasible up to $|\mathcal{X}| \sim 10^3$.

Worked example on Fig. 3. Consider a routing/navigation user heading east: the service penalizes *westward* reporting errors much more heavily than eastward (a westward-skewed report misroutes the vehicle). Model with the asymmetric, direction-dependent loss

$$c_{x,y} \;=\; 5\max\big(0,\, \mathrm{lon}(x) - \mathrm{lon}(y)\big) \;+\; 2\,\big|\mathrm{lat}(x) - \mathrm{lat}(y)\big|,$$

i.e., a unit westward shift costs 5, eastward 0, latitude 2. With $\mathcal{X}$=$\mathcal{Y}$={A, .., E}, $d_{\mathcal{X}}$=d_g, uniform π, ε=$\ln 2$: solving the LP at x=D gives $\Pr[E]$=0.833, $\Pr[B]$=0.167 (other entries 0; optimal expected

loss ≈ 0.717). B (north neighbor) and E (east neighbor) both lie at *graph distance* 1 from D, so any HDM/NHDM is forced by symmetry under d_g to assign them equal probability ($\Pr[B]=\Pr[E]=0.195$ for EM at $\varepsilon=\ln 2$); the LP exploits the cost asymmetry $c(D,B)=2$ vs. $c(D,E)=0$ to push mass eastward as far as the mDP constraints allow — the maneuver HDM/NHDM *cannot* make with a closed-form distance-only kernel.

Privacy/utility. The LP solution is provably optimal for the chosen $(\pi, c, d_{\mathcal{X}}, \varepsilon)$ [110]; equivalently, OPM achieves the lowest expected loss among all $(\varepsilon, d_{\mathcal{X}})$-mDP mechanisms. Its bottleneck is the cubic constraint count, limiting direct deployment to $|\mathcal{X}| \lesssim 10^3$ [3].

LP+EM hybrid (Imola et al., UAI 2022). [3] cuts the LP size by *fixing* far-away perturbation probabilities to a closed-form EM kernel and only optimizing the near-neighbor entries. The result is an NHDM kernel for distant pairs and an LP-tuned tail—structurally a non-homogeneous distance mechanism whose parameters are computed by an OPM, which we discuss as a canonical hybrid in Sec. 3.4. Other strategies attacking OPM's scalability include hierarchical/multi-step indices [78] and dataset-partition with Benders or Dantzig–Wolfe decomposition [32, 111] (large-scale LP techniques: Benders splits the LP via cut generation on complicating variables, while Dantzig–Wolfe uses column generation to expand the master problem from sub-LPs).

Beyond the Bordenabe LP anchor, the OPM literature populates five orthogonal degrees of freedom of the LP recipe ($\min_{\mathbf{Z}} \pi^\top(\mathbf{Z}\odot c)\mathbf{1}$ subject to mDP constraints), detailed in Table 9. *(A′) The cost matrix* $c_{x,y}$: replacing geometric distance with task-specific objectives — task-completion quality [120], MINLP travel-distance for crowdsourcing matching [115, 116], multi-criteria objectives that combine conditional entropy with worst-case quality loss [135], endpoint-shift utility [124], ride-on-demand pricing-aware loss [2], group-based traffic-flow cost [122], and multi-point "Geo-Ind masking" that produces an obfuscation *set* rather than a single point [113]. *(B′) Constraint structure and scalability*: the cubic constraint count [110] is attacked through hierarchically well-separated trees [78], Benders decomposition with dataset partitioning [111], Dantzig–Wolfe column generation on road networks [121], peer-set constraint pruning for time-critical settings [123], graph-based CORGI approximations that decouple server-side optimization from user-side customization [114], and shortest-path metric reformulation on road networks [32]. *(C′) Cross-family hybrids*: the canonical LP+EM hybrid [3] fixes far-away entries to a closed-form EM kernel and only optimizes near-neighbor entries (also discussed as the canonical hybrid in Sec. 3.4); Bayesian remapping [74] composes a closed-form HDM with an LP-derived post-processor. *(D′) Alternative optimization paradigms*: when the LP itself is too rigid or the domain is too high-dimensional, the field turns to Contract Theory under information asymmetry for cooperative localization [89], Stackelberg / bilevel games for accuracy–privacy balance [125], deep reinforcement learning (A3C) for 3D-spatial task allocation [126], and swarm/heuristic optimizers for ride-on-demand [2]. *(E′) Theoretical foundations and system integration*: dual-problem proofs of HDM optimality [19], metric-geometry characterizations of the privacy-utility frontier [129], rigorous re-analysis of PIVE that identified privacy-proof flaws and proposed corrections [96], IoT proxy-gateway selection under DP for energy-efficient smart homes [127], and traffic-flow-aware fake-trajectory selection [128]. The shared theme: OPM is the only family that can express *anisotropic, direction-dependent* utility cleanly, but at a cost in both LP scale and solver flexibility, motivating the hybrid and alternative-paradigm research that dominates the recent literature.

Takeaway. OPM treat the perturbation kernel as a constrained-optimization variable rather than a closed-form distance kernel: probabilities $z_{x,y}$ minimize an arbitrary task-specific loss $c_{x,y}$ subject to mDP. This is the only family that can express *anisotropic, direction-dependent* utility cleanly — as the worked example above shows, OPM gave $\Pr[\text{east}]/\Pr[\text{north}] \approx 5{:}1$ at $x = D$, while HDM and NHDM are forced to 1:1 by symmetry under d_g. The price is computational: the canonical Bordenabe LP has $\Theta(|\mathcal{X}|^2|\mathcal{Y}|)$ privacy constraints, limiting direct deployment to small/medium domains. Three orthogonal scalability strategies dominate the follow-up literature — hierarchical/multi-step decomposition [78], dataset partition with Benders/Dantzig–Wolfe [32, 111], and LP+EM hybrids [3] — but none yet achieves "OPM at the scale of HDM" on continuous high-dimensional inputs (open problem in Sec. 5.2).

3.4 Comparison

In comparing the three categories of mechanisms, each exhibits distinct strengths and limitations in terms of *adaptability*, *utility preservation*, and *time efficiency*; Table 10 in App. A provides a comparative overview of major works. Across the 174 surveyed references — 84 employing HDM, 35 NHDM, and 36 OPM (the rest hybrid or non-perturbation) — a clear progression emerges from early Laplace-based HDM to neighborhood-aware NHDM and to optimization-driven OPM, mirroring both the widening application range and advances in scalability.

Adaptability and utility. HDM and NHDM both tie probability mass to $d(x, y)$ alone, so two outputs at the same metric distance receive the same probability regardless of how task utility shapes the loss. NHDM [142] adds value over HDM mainly when outputs must lie in a discrete or structured set (graphs, vocabularies, road points): a noisy embedding from a homogeneous Laplace mechanism is unlikely to correspond to any valid word, and post-hoc nearest-neighbor projection [26, 28] treats the embedding space as non-sensitive, potentially overlooking the privacy leakage of the projection itself. Only OPM expresses *anisotropic, direction-dependent* utility cleanly, by encoding a task-specific cost matrix $c_{x,y}$ and directly minimizing expected loss under mDP — as the running-example LP at $x = D$ shows, OPM gave $\Pr[\text{east}]:\Pr[\text{north}] \approx 5:1$ where HDM/NHDM are forced to $1:1$ by symmetry under d_g.

Time efficiency. HDM and NHDM are scalable: planar Laplace [5] samples in $O(1)$, the EM construction of [92] in $O(n)$ per release, with structured candidate sets inflating to $O(n^2)$ in the worst case [60, 143]. OPM is roughly cubic ($\Theta(|\mathcal{X}|^2|\mathcal{Y}|)$ constraints), limiting direct deployment to $|\mathcal{X}| \sim 10^3$ [3, 78]; the recent decomposition wave (HST, Dantzig–Wolfe, Benders) closes much of the gap but does not fully eliminate it [32, 111, 140, 141].

Hybrids and the utility-proxy issue. The HDM/NHDM/OPM partition is a continuum, and several influential designs straddle two categories. Three canonical hybrid patterns recur: *LP-tuned*

The detailed classification can be found on GitHub at https://github.com/Kerwinxxp/metricDP.

EM [3] keeps the closed-form EM kernel but sets parameters via OPM-style LP, fixing far-away entries to a weighted exponential and only optimizing near neighbors; *Bayesian-remapped Laplace* [74] pairs an HDM kernel with an OPM-style remap to the highest-utility output consistent with a prior; *constrained-EM variants* (e.g., truncated EM [142]) keep EM's exponential form but add explicit utility constraints — all flagged with multiple checkmarks plus a $^{\mathrm{H}}$ marker in Table 10. The reason hybrids exist is the same reason geometric utility proxies are loose: pure HDM/NHDM kernels are *radial and direction-agnostic*, while most downstream task utilities are direction-sensitive — comparable expected-ℓ_2 values can yield substantially different POI-retrieval precision [5], classification F1 over noisy embeddings [20], or graph community-recovery accuracy. We therefore recommend, as a benchmarking guideline tied to Sec. 5, that future mDP work *report task-driven utility alongside any geometric proxy* and match mechanisms at fixed task-utility rather than at fixed ε or fixed geometric distortion. Table 10 annotates entries with $^{\mathrm{G}}$ (geometric distortion) or $^{\mathrm{T}}$ (task-driven utility) accordingly.

Takeaway. Across HDM/NHDM/OPM, the practical choice is governed by the same three trade-offs: *adaptability* (HDM is most rigid, OPM most flexible), *utility-faithfulness* (HDM/NHDM tie probability to $d(x, y)$ alone; only OPM expresses anisotropic loss), and *scalability* (HDM and NHDM are $O(1)$–$O(n^2)$ per release while OPM is roughly cubic before decomposition). Hybrid mechanisms (LP-tuned EM kernels, Bayesian-remapped Laplace, constrained-EM variants) populate the continuum between families and are often the most practical choice in deployment. A largely orthogonal axis — under-recognized in current evaluations — is whether the reported *utility metric* is a geometric proxy or a downstream-task measure: future mDP studies should match mechanisms at fixed *task-driven utility*, not at fixed ε or fixed geometric distortion (see Sec. 5).

4 Applications

Since its introduction in 2013, mDP has proven to be a versatile and effective framework across diverse application domains, largely due to its ability to tailor privacy guarantees to domain-specific distance metrics. As shown in Figure 1, research activity has grown steadily, with location-based applications remaining the dominant focus, while increasing interest in text, image, and voice data highlights mDP's expanding reach and influence.

4.1 Application Domains

Location privacy. The most extensively studied mDP domain. Two structurally distinct sub-domains have emerged: *mobile spatial crowdsourcing (MSC)* pairs ℓ_2-Geo-Ind on workers with task assignment, while *intelligent transportation systems (ITS)* replaces ℓ_2 with road-graph metrics for routing-faithful obfuscation, often supplemented by validity guarantees that exclude implausible outputs (Perturbation-Hidden [102]). Recent infrastructure-scale deployments — blockchain-coordinated truncated Geo-Ind for indoor paging [63] and electric-vehicle charging-station recommendation under $d_{\mathcal{X}}$-privacy [49] — show that mDP has matured well beyond academic LBS settings.

Text data privacy. The decisive structural shift here is from the *continuous* embedding space to the *discrete* vocabulary: HDM applied to embedding ℓ_2 [28] or syntactic distance [84] samples points that may not correspond to any valid word, motivating either nearest-neighbor projection back to the vocabulary [26] (which itself constitutes additional privacy leakage if the embedding-to-token map is treated as non-sensitive) or NHDM mechanisms operating directly over the vocabulary [103, 104]. Recent advances calibrate noise to local embedding density (Truncated EM [142]) or simplify via dimension reduction (1-Diffractor [145]); Faustini *et al.* [51] explicitly bridges mDP and the privacy-preserving language-model literature. We note that several recent works applying local Laplace noise to LLM prompts or embeddings (typically labeled "DP for LLM") are *de facto* mDP mechanisms despite the label — the noise is calibrated to a learned-embedding metric rather than to Hamming-style adjacency on a discrete alphabet, motivating the LLM service stack below.

LLM service stack. The LLM era is converting the "de-facto mDP" observation above into a concrete deployment stack along the inference–training pipeline.

(a) Token-level prompt sanitization. Pr$\epsilon\epsilon$mpt [176] sanitizes value-dependent tokens (age, salary) under metric LDP and uses format-preserving encryption for format-only tokens (SSN, credit cards), drawing an explicit cryptography-vs-mDP boundary. The $d_{\mathcal{X}}$-Stencil mechanism [177] couples semantic distance with local context vectors to attain 2ε-$d_{\mathcal{X}}$-privacy, mitigating context leakage. STAMP [180] introduces a *polar mechanism* that perturbs only the *direction* of token embeddings on the unit sphere and allocates budget by task importance — a direct instantiation of our task-aware utility recommendation in the LLM stack.

(b) Embedding- and span-level mDP at inference. Split-and-Denoise (SnD) [178] runs the embedding layer client-side under $d_{\mathcal{X}}$-privacy with a server-side denoiser. PrivacyRestore [179] excises privacy spans and applies $d_{\mathcal{X}}$-privacy only to a learned meta-restoration vector, avoiding the linear budget growth of span-by-span mDP. CMAG [183] adapts the analytic-Gaussian mechanism to a Mahalanobis metric on sentence embeddings with locally density-adaptive noise — a sentence-level analogue of the elastic/PRIVIC line (Sec. 3.2).

(c) Training and fine-tuning. RAPT [184] privatizes inputs via text-to-text mLDP before PEFT, with an auxiliary token-reconstruction objective. To our knowledge it is the closest analogue of a metric-aware DP-SGD, but acts on PEFT *inputs* rather than *gradients* (Sec. 5.6).

(d) LLM-as-adversary against mDP. LLMs are also new adversaries against word-level mDP outputs. Pang *et al.* [182] report black/white-box reconstruction attacks recovering 70–94% of original tokens from SanText/CusText. Meisenbacher *et al.* [181] characterize this as *contextual vulnerability*: independent per-word perturbation leaves syntactic/co-occurrence cues for an LLM to exploit, though the same LLM can also serve as a defensive paraphraser. mDP-sanitized text must now be evaluated against contextually-aware LLM adversaries.

Image, voice, graph, and FL/IoT. All four moved from raw signal to feature-space perturbation, but with very different metric structures: image privacy uses perceptual or GAN-latent metrics [8, 130]; voice obfuscation uses angular distance in x-vector space; graph privacy combines node-edit distance with Lipschitz extensions for vector-valued statistics like degree distributions and subgraph counts; and federated/IoT settings adapt d-privacy to group-level guarantees and energy-aware proxy selection. The unifying pattern across all domains: mDP's value-add is the choice of *metric* to encode each domain's notion of similarity, with the HDM/NHDM/OPM mechanism choice following the structural properties of the input — HDM for continuous coordinates and embeddings, NHDM for discrete vocabularies and road points, OPM where task utility is anisotropic. A growing trend is to combine $d_{\mathcal{X}}$-privacy with system-level layers (blockchain, federated rounds, shuffle aggregation) and infrastructure-scale deployments rather than rely on it in isolation.

4.2 Context-Aware mDP Mechanisms

Context-aware mDP recognizes that the right ε *and* the right $d_{\mathcal{X}}$ are user/scene-dependent: mechanisms in this family parameterize $\varepsilon(\cdot)$ or $d_{\mathcal{X}}(\cdot)$ by mobility history [23], query rank [82], area density [92], or community covariance [37]. Three flavors recur: *policy-driven* (PGLP [33], Blowfish [31]), where users specify protected pairs explicitly; *geometry-driven* (Geo-Ellipse [37], semantic Geo-Ind, the tree-structured customizable framework of Pappachan *et al.* [114]), where covariance or land-use estimates reshape the metric; and *rank-driven* (PIVE [95], top-k schemes [82]), where downstream task importance modulates the budget. Recent extensions integrate group-wise k-anonymity with DP [174] and game-theoretic configurations [132]. The principled cost of personalization is non-uniform composition (Sec. 5.3): when $d_{\mathcal{X}}$ varies across releases, sequential bounds degrade by Lipschitz scaling rather than direct addition, and user-specific metrics complicate the privacy-budget audit because adversaries may exploit the metric choice itself.

4.3 Dataset

The evaluation of mDP mechanisms relies on a diverse set of public and synthetic datasets summarized in Table 11. Reflecting Fig. 1, location/mobility datasets dominate (trajectory benchmarks like GeoLife, T-Drive, Cabspotting; LBSN check-ins like Gowalla, Brightkite, Foursquare), supporting work on continuous-trace privacy, predictive mechanisms, and policy-driven frameworks. Text corpora form the second wave (IMDb, PAN/Enron, movie reviews) for sentiment masking, author obfuscation, and embedding fine-tuning. Image (EMNIST), tabular census (US ACS, US Cities), and road-network resources (OpenStreetMap, Google Places [171]) cover the remaining settings; synthetic data provides reproducible ground truth but cannot substitute for real distribution shift in continuous/streaming settings. A recurring gap across all these benchmarks is the absence of standardized *task-driven* measurements at matched ε: most studies report geometric distortion (expected ℓ_2 error, Wasserstein distance) rather than POI recall, classification F1, or community-recovery accuracy — a measurement convention that, as Sec. 3.4 argues, can hide real differences between mechanisms; we list standardized task-driven benchmarks as a community priority in Sec. 5.8.

Takeaway. Across location, text/LLM, image, voice, graph, and FL/IoT (Table 4), the application landscape exhibits a single structural pattern: the choice of *metric* encodes each domain's notion of similarity, and the HDM/NHDM/OPM mechanism choice then follows mechanically from the structural properties of the input — HDM for continuous coordinates and embeddings, NHDM for discrete vocabularies and road points, OPM where task utility is anisotropic or constraint-driven. This pattern is what links Sec. 4 to Sec. 5: the open problems we develop next (composition, amplification, local-model statistical inference, and mDP for deep learning) are downstream consequences of the same structural tension that organizes Table 4 — namely, that each domain's mDP deployment is bottlenecked either by the direction-agnosticism of HDM/NHDM or by the deployment cost of OPM.

5 Open Challenges and Research Agenda

The application landscape of Sec. 4 shows that mDP is already deployed across location, text/LLM, image, voice, graph, and FL/IoT settings. Yet four cross-cutting issues remain underdeveloped: classical ε-DP machinery — advanced composition, subsampling amplification, matrix-inversion estimators, gradient-clipping accountants — assumes bounded sensitivity under fixed adjacency, whereas mDP uses domain-specific, often unbounded and heterogeneous metrics. Closing these gaps is necessary for mDP to become a deployable, mature privacy framework.

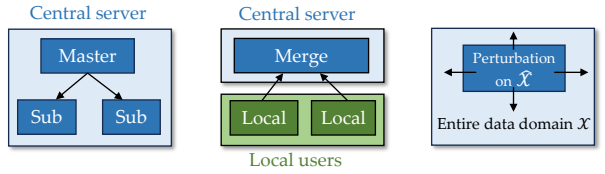
We organize the open problems along three dimensions: (i) *scalable and deployable mechanisms*; (ii) *theoretical foundations*; and (iii) *metric design and evaluation*. The key technical frontiers — composition, amplification, Iterative Bayesian Update (IBU)-based estimation, and mDP-SGD — are downstream of the OPM-vs-HDM-scale tension revealed by the taxonomy.

5.1 mDP Mechanism Selection

Each of the three mechanism families occupies a different point in the privacy–utility–efficiency design space (Sec. 3): HDM offers simplicity and strong theoretical guarantees but underfits heterogeneous utility landscapes; NHDM adapts to local density but requires careful candidate-set and metric design; OPM tailors perturbations to arbitrary task-induced loss surfaces at substantial computational cost. Existing work provides only problem-specific heuristics — e.g., distance-based noise when domains are large or losses appear radial [5], optimization when domains are small and losses are clearly non-radial [141] — with thresholds and adequacy scores tuned per dataset. Systematically deciding, for a new application and metric, *which* family (or hybrid) to deploy remains open. The challenge is to design *meta-selection* procedures that, given a metric, an input prior, and a task-utility profile, can identify the underlying loss regime (radial / directional / heterogeneous), anticipate when a simple HDM/NHDM is near-optimal versus when OPM is warranted, and accomplish this under realistic constraints on computation and on the privacy cost of any data used for calibration. Progress likely requires both theoretical results that characterize when each family admits provable approximation guarantees relative to an oracle optimum, and practical diagnostic tests runnable on limited synthetic or public data without further privacy leakage.

Table 4: Cross-domain systematization of mDP designs. For each domain in Sec. 4.1, we list typical metrics, the dominant mechanism family under our HDM/NHDM/OPM taxonomy, the distinctive challenge driving mechanism choice, and representative references.

Domain	Typical metric(s)	Dominant family	Distinctive challenge	Representative refs
Location (LBS, MSC, ITS)	Euclidean ℓ_2; shortest-path on road graph	HDM (planar Laplace); OPM (LP-based)	Output validity; road-graph constraints; outlier disclosure under repeated release	[5, 32, 49]
Text / NLP	Word Mover's distance; embedding ℓ_2	NHDM (EM over vocabulary)	Off-vocabulary samples; high-dim utility collapse; semantic coherence	[20, 105, 142]
LLM service stack	Learned embedding ℓ_2 / cosine; token-context vectors	NHDM + Hybrid (EM kernel + reconstruction)	Contextual vulnerability; LLM-as-adversary; gradient-stage gap (no mDP-SGD analogue)	[176, 179, 182, 184]
Image	Perceptual / GAN-latent; SVD-space; composite (ℓ_2+JSD)	HDM (in feature space)	Manifold validity in feature space; feature-extractor privacy	[9, 146]
Voice	Angular distance in x-vector space	HDM	Spoofing under speaker re-identification	[7]
Graph	Node-edit distance; structural distance	Lipschitz extension (HDM-flavored)	Vector-valued statistics; high node-level sensitivity	[30, 107]
FL / IoT	d-privacy at group level	HDM (FL); OPM (IoT)	Cross-device coordination; energy-budget constraints	[91, 127]



(a) Decomposition (b) Distributed optimization (c) Extension

Figure 5: Comparison of scalable optimization approaches.

5.2 Scalable Optimization

Scalability remains the primary obstacle to deploying utility-preserving mDP mechanisms in practice: OPM offers strong utility and flexible metric support but incurs cubic constraint counts that limit current methods to centralized servers and small-medium domains (Fig. 5(a)). Existing decomposition work [111, 140] (grouping locations, exploiting graph structure, separating feature blocks) improves scalability but does not eliminate the centralization bottleneck.

Two complementary directions stand out. *Distributed/locally-optimized mechanisms* (Fig. 5(b)): partition the metric space (Voronoi cells, clusters, tiles), solve mDP locally per region, then merge via consistency constraints on region boundaries; this fits naturally into distributed deployment with only lightweight boundary exchange. *Extension algorithms* (Fig. 5(c)): start from a mechanism optimized on a core anchor set $\hat{\mathcal{X}}$ and extend to a larger or fine-grained $\mathcal{X}$. A recent example [112] partitions an N-dimensional input domain into cells, optimizes perturbations on cell-corner anchors via an *Anchor Perturbation Optimization* (APO) problem, and extends to non-anchor points through coordinate-wise log-convex (weighted geometric-mean) interpolation that preserves global Lipschitz/mDP constraints. Such extension schemes retain rigorous guarantees on the optimized core while incurring only an explicitly quantified slack outside it. Open directions: hierarchical mDP mechanism design with provable approximation, relaxed mDP guarantees that preserve adversarial interpretability while reducing computation, and reusable solver libraries for common metrics and domain decompositions.

5.3 Composition in mDP

Composition is a central issue in mDP because practical systems often release multiple perturbed outputs, possibly under different metrics or across different components of the input. Unlike classical DP, however, the mDP literature does not yet provide a unified composition theory. We summarize the current status in Table 5.

The most direct case is *sequential composition* under the same metric. Similar to classical ε-DP, if two releases satisfy $(\varepsilon_1, d_{\mathcal{X}})$-mDP and $(\varepsilon_2, d_{\mathcal{X}})$-mDP, respectively, then their joint release satisfies $(\varepsilon_1+\varepsilon_2, d_{\mathcal{X}})$-mDP [4, 19]. This result provides the basic linear accounting rule for repeated releases over the same metric space.

A more subtle case arises when different stages use different metrics, such as an ℓ_2 distance in one stage and an embedding-based semantic distance in another. When the metrics are comparable, composition can be handled through a Lipschitz conversion. Specifically, if $d_{\mathcal{X}_2}(x, x') \le L \cdot d_{\mathcal{X}_1}(x, x')$ for all x, x', then composing an $(\varepsilon_1, d_{\mathcal{X}_1})$-mDP release with an $(\varepsilon_2, d_{\mathcal{X}_2})$-mDP release yields $(\varepsilon_1 + L\varepsilon_2, d_{\mathcal{X}_1})$-mDP. This bound is useful as a general worst-case rule, but it can be loose when the Lipschitz constant is large, infinite, or dominated by rare pairs, as may occur for shortest-path metrics or learned semantic metrics.

In contrast, *advanced composition* for mDP is much less developed. Classical $\sqrt{k \ln(1/\delta)}$-style composition relies on concentration of bounded per-step privacy losses under a fixed neighboring relation. In mDP, the privacy-loss bound scales with distance, namely $\varepsilon d_{\mathcal{X}}(x, x')$, which may be unbounded on continuous or large-diameter domains. Therefore, without additional assumptions such as bounded diameter, localized adjacency, or mechanism-specific tail bounds, classical advanced-composition arguments do not directly yield a general mDP accountant. Existing progress is mostly mechanism-specific or divergence-based, including analyses for planar Laplace and geometric mechanisms; a unified advanced-composition theorem remains open.

A complementary line of work studies *dimension-wise composition* for ℓ_p-mDP [112]. Consider a domain $\mathcal{X} \subseteq \mathbb{R}^N$ equipped with the ℓ_p metric. If changing coordinate ℓ alone satisfies an (ε_ℓ, d_1)-mDP guarantee, then the mechanism satisfies global (ε, d_p)-mDP

Table 5: Status of composition results under mDP.

	Same metric $d_{\mathcal{X}}$	**Comparable metrics $d_{\mathcal{X}_1}, d_{\mathcal{X}_2}$**
Sequential / linear	Tight: $\varepsilon_1 + \varepsilon_2$ [4, 19]	Lipschitz conversion: $\varepsilon_1 + L\varepsilon_2$
Advanced ($\sqrt{k}$)-style	Partial: divergence-based or mechanism-specific bounds	Largely open
Product / dimension-wise	$\sum_\ell \varepsilon_\ell^{p/(p-1)} \le \varepsilon^{p/(p-1)}$ for $p > 1$ [112]	

whenever

$$\sum_{\ell=1}^{N} \varepsilon_\ell^{p/(p-1)} \le \varepsilon^{p/(p-1)} \ (p > 1), \quad \max_\ell \varepsilon_\ell \le \varepsilon \ (p = 1). \tag{12}$$

The proof decomposes a multi-coordinate displacement into an axis-aligned path and applies Hölder's inequality. This result gives a metric-aware alternative to standard linear accounting. For example, under ℓ_2, equal allocation gives $\varepsilon_\ell = \varepsilon/\sqrt{N}$, while asymmetric allocations can exploit directional sensitivity in non-isotropic data. In the limit $p \to \infty$, the condition recovers the standard sequential-style rule $\sum_\ell \varepsilon_\ell \le \varepsilon$.

Overall, composition in mDP is well understood for linear composition under a fixed metric, partially understood for comparable heterogeneous metrics and product metrics, and largely open for advanced composition. Important open questions include developing tight advanced-composition bounds for unbounded or large-diameter metrics, establishing sharper guarantees when the metric comparison constant L is learned or data-dependent, and integrating metric-aware accounting with amplification techniques such as shuffling and subsampling.

5.4 Privacy Amplification: Shuffling and Subsampling

In classical ε-DP, two general-purpose amplifications turn local-model into central-model-quality guarantees: *shuffle amplification* [47, 48] via a trusted shuffler ($\widetilde{O}(\varepsilon\sqrt{\log(1/\delta)/n})$ boost) and *subsampling amplification* ($\varepsilon \to \log(1 + q(e^\varepsilon - 1))$). Both transfer to mDP only with friction.

Shuffle amplification. The first systematic study [46] surfaces obstacles specific to mDP. *Outlier identifiability*: metric-faithful mechanisms (Geometric, planar Laplace, Truncated EM) concentrate output mass near the true input, so an outlier survives the permutation as recognizable in the shuffled multi-set — naive shuffling yields no amplification. The fix encodes each truncated noisy output as a unary bit vector so the shuffled bit-pile is privacy-equivalent to a bit-sum, but requires new analysis through the previously-unstudied *Symmetric Generalized Discrete Laplace (SGDL)* distribution and an unavoidable additive δ slack from truncation — pure ε-mDP shuffle amplification remains open. The SGDL- vs Geo-Shuffle constructions further encode an mDP-specific *amplification–robustness* trade-off: maximizing central-model utility makes local-model privacy collapse if the shuffler is compromised, whereas graceful degradation costs amplification.

Subsampling amplification. Classical subsampling relies on a fixed per-pair ε. Under mDP the per-pair guarantee scales with $d_{\mathcal{X}}(x, x')$, so amplification depends on the population's distance distribution and its interaction with the sampling rate q in ways that have not been characterized. To our knowledge, no published work systematically analyzes mDP subsampling — a notable gap given how heavily classical privacy accountants rely on it. Several questions therefore remain open: whether pure ε-mDP shuffle amplification can be achieved without the δ slack introduced by truncation; how to extend shuffle amplification beyond ℓ_p metrics to shortest-path metrics on road networks and to learned or context-adaptive metrics (cf. Sec. 5.7); and whether a metric-aware subsampling accountant can be developed that composes cleanly with the dimension-wise rule of Sec. 5.3.

5.5 Local-Model Statistical Utility: IBU and Beyond

Real deployments of locally-mDP collection feed downstream analytics that need population-level estimates. The default LDP estimator is matrix-inversion (Inv) [53, 54]: with channel matrix $\mathbf{Q}_{xy} = \Pr[\mathcal{M}(x)=y]$ and empirical output histogram $\hat{q}$, report $\hat{\pi} = \mathbf{Q}^{-1}\hat{q}$. Inv works well when $\mathbf{Q}$ is approximately uniform off-diagonal (e.g., k-RR). *Metric-faithful mechanisms violate this by design*: planar Laplace, Geometric, and EM with $u = -d_{\mathcal{X}}$ all sharply peak $\mathbf{Q}$ on the diagonal, $\kappa(\mathbf{Q})$ grows essentially without bound as discretization is refined, $\mathbf{Q}^{-1}$ amplifies noise, and Inv produces high-variance estimates with frequent negative entries that must be clipped [55, 56].

The *Iterative Bayesian Update* (IBU) [52, 55] avoids explicit inversion by formulating recovery as maximum-likelihood estimation under $\mathbf{Q}$ and solving it via EM. Three observations are especially relevant here: for near-uniform mechanisms, IBU reduces to Inv; for sharply peaked, metric-faithful mechanisms—the regime most relevant to mDP—IBU is more numerically stable and has been shown to achieve lower MSE in the studied settings [56]; and, as a post-processing step, it incurs no additional privacy cost. Recent work further re-derives IBU's convergence directly from EM and establishes statistical consistency [57]. A practical takeaway is therefore that locally deployed mDP mechanisms should typically pair metric-faithful perturbation with IBU-style reconstruction rather than naive inversion.

Several questions remain open: finite-sample guarantees under common metric mechanisms; extensions beyond histogram estimation to ranking, density estimation, regression, and model fitting; and more scalable implementations for large domains. In particular, while each IBU iteration is quadratic in the domain size, the total runtime depends on the number of iterations required for convergence, which is currently not well characterized.

5.6 mDP and Deep Learning: Opportunities and Obstacles

Given the dominance of DP-SGD in private deep learning, how does mDP fit? The picture is asymmetric: mDP is widely deployed at the *input-perturbation* stage — word/embedding-level planar Laplace for NLP [20, 51, 142], coordinate-level planar Laplace for geolocation features — but largely absent in iterative training. The metric structure is exploited once before any model interaction; the noise is frozen and the rest of the pipeline proceeds without mDP machinery. The closest extant analogue is RAPT [184], which privatizes inputs via text-to-text mLDP before parameter-efficient

fine-tuning and trains an auxiliary token-reconstruction head to recover utility; the noise, however, is added once to the inputs rather than to gradients across iterations. We are not aware of an mDP analogue of DP-SGD that injects metric-calibrated noise into gradient updates.

The principal obstacle is composition (Sec. 5.3): DP-SGD's Rényi-DP / moments accountant presupposes bounded per-step sensitivity, but mDP's per-step loss scales with the unbounded $d_{\mathcal{X}}$, so T identical mDP gradient releases yield an $O(T)$ bound rather than $O(\sqrt{T})$. A second obstacle is benchmarking: comparing DP-SGD and an mDP analogue via raw ε values is misleading because DP's adjacency is a unit-cost Hamming relation while mDP's is graded by a real-valued metric, the same numerical ε encodes different operational guarantees. The cleaner comparison is at *matched adversarial advantage* on a downstream task. Three concrete open problems follow from this picture: developing tight advanced composition under an unbounded $d_{\mathcal{X}}$; identifying metric structures on the *gradient* space (rather than the input space) that admit informative mDP guarantees over iterative training; and designing mDP-compatible accountants that interface cleanly with existing DL pipelines.

5.7 Learnable and Context-Aware Metrics

The metric $d_{\mathcal{X}}$ is the primary carrier of semantics and adversarial assumptions in mDP — it encodes which inputs are "locally indistinguishable," which directions incur utility loss, and which attacks the guarantee covers. Yet most existing work adopts off-the-shelf metrics (Euclidean, Manhattan, Wasserstein, angular, shortest-path) without a principled way to select or adapt them to real data and tasks.

A key direction is *learnable, context-aware metrics* that adapt to data semantics while satisfying mDP's structural requirements. *Metric learning under mDP constraints*: learn a parametric distance (Mahalanobis-type or embedding-based $d_{\mathcal{X}}(x,x') = \|f_\theta(x) - f_\theta(x')\|$) subject to validity (metric/pseudometric axioms), Lipschitz bounds for noise calibration, and basic interpretability. Concrete instances per domain: in location privacy, incorporating travel time and population density rather than raw Euclidean distance; in text, pushing apart pairs that differ in sensitive attributes while keeping benign paraphrases close; in tabular data, weighting feature groups (clinical vs. demographic) by expert-specified risk; for images, learning perceptual or task-specific distances aligned with the changes that actually matter. *Privacy-aware metrics from richer formalisms*: posterior-leakage metrics (mPL-style), probabilistic mDP variants, or PAC-style guarantees [119] where protection is formulated in terms of how much posteriors or leakage measures can change for most input pairs or with high probability. Metric learning can then be guided by minimizing empirical violation rates, yielding a feedback loop in which leakage-based objectives supervise the learning of $d_{\mathcal{X}}$ and the learned metric supports stronger probabilistic mDP-style guarantees.

5.8 Benchmarks and Geometry-Aware Evaluation

As Sec. 4 highlights, current mDP evaluations are highly fragmented: different datasets, metrics, threat models, and utility measures make cross-paper comparison difficult. Unlike DP-SGD in machine learning [117] or standard LBS benchmarks in location privacy [118], mDP still lacks a shared experimental foundation or even a canonical set of tasks and metrics.

A mature mDP ecosystem requires four components. *(1) Standardized benchmarks* spanning location, graph, text, embedding, and multi-modal domains, with explicit admissible metrics and canonical tasks for cross-paper utility comparison. *(2) Reference implementations of mechanisms and attackers*: vetted libraries for Laplace, EM, optimization-based designs, remapping/post-processing, adaptive variants, and standard attacker models. *(3) Geometry-aware evaluation tools* that visualize metric balls, local sensitivity, and mechanism-induced distortion, and report privacy-utility trade-offs across distance scales or regions. *(4) Reporting and reproducibility standards*: papers should report the exact metric, threat model, ε range, utility tasks, runtime/resources, and post-processing. A unified "mDP-Bench" suite would provide a practical playbook for selecting and tuning mDP mechanisms.

6 Conclusions

Over the past decade, mDP has matured into a central generalization of classical DP, providing distance-aware indistinguishability guarantees that align with structured and continuous input domains in location-based services, text and language data, and multimedia. This SoK systematized the mDP landscape through the HDM/NHDM/OPM taxonomy and its hybrids, and connected their utility trade-offs, application patterns, and scalability considerations.

The structural punchline. Looking across ten years of literature through the HDM/NHDM/OPM lens reveals a pattern not visible from any single subfield. The continuous-symmetric case is essentially solved: planar Laplace and its ℓ_p relatives are MSE-optimal HDMs and scale to $O(1)$ per release. The discrete-symmetric case is well-handled: EM and its elastic / Blahut–Arimoto refinements deliver near-optimal NHDM utility on finite candidate sets. The remaining open problems concentrate in the OPM corner, where task utility is *anisotropic, direction-dependent, or constraint-driven.* Every domain in Table 4 confirms this: applications dominated by HDM/NHDM are bottlenecked by direction-agnosticism, while applications demanding OPM are bottlenecked by its $\Theta(|\mathcal{X}|^2|\mathcal{Y}|)$ LP. The decomposition wave (HST, Dantzig–Wolfe, Benders) and LP+EM hybrids are not independent research threads; they are angles of attack on the same problem: *how to recover OPM's expressive power at HDM's deployment cost.*

On this basis, we surface four under-developed frontiers — *composition* (Sec. 5.3), *amplification via shuffling/subsampling* (Sec. 5.4), *local-model statistical inference* (Sec. 5.5), and *mDP for deep learning* (Sec. 5.6) — each requiring OPM-like flexibility with HDM-like analytical tractability.

Looking forward, the productive question is no longer *"which family is best?"* — the taxonomy answers that per setting — but *how far OPM can be pushed toward HDM's scale*, and dually, *how much anisotropic expressiveness can be smuggled into HDM/NHDM kernels via post-processing, hybridization, and metric design.* Progress on either front will translate a decade of theoretical gains into deployable privacy for large-scale learning, real-time systems, and the LLM service stack.

Acknowledgments

The authors used generative AI-based tools to revise the text, improve flow and correct any typos, grammatical errors, and awkward phrasing. This research was partially supported by U.S. NSF grants CNS-2136948, CNS-2313866, and JST PRESTO JPMJPR23P5.

References

[1] C. Dwork, A. Roth *et al.*, "The algorithmic foundations of differential privacy," *Foundations and Trends® in Theoretical Computer Science*, vol. 9, no. 3–4, pp. 211–407, 2014.

[2] Z. Zheng, Z. Li, S. Long, S. Guo, C. Chen, and K. Xu, "Pricing Utility vs. location privacy: a differentially private data sharing framework for ride-on-demand services," *IEEE Transactions on Dependable and Secure Computing*, 2025.

[3] J. Imola, S. Kasiviswanathan, S. White, A. Aggarwal, and N. Teissier, "Balancing utility and scalability in metric differential privacy," in *Uncertainty in Artificial Intelligence*. PMLR, 2022, pp. 885–894.

[4] K. Chatzikokolakis, M. E. Andrés, N. E. Bordenabe, and C. Palamidessi, "Broadening the scope of differential privacy using metrics," in *international symposium on privacy enhancing technologies symposium*. Springer, 2013, pp. 82–102.

[5] M. E. Andrés, N. E. Bordenabe, K. Chatzikokolakis, and C. Palamidessi, "Geo-indistinguishability: Differential privacy for location-based systems," in *Proceedings of the 2013 ACM SIGSAC conference on Computer & communications security*, 2013, pp. 901–914.

[6] O. Feyisetan and S. Kasiviswanathan, "Private release of text embedding vectors," in *Proceedings of the First Workshop on Trustworthy Natural Language Processing*, 2021, pp. 15–27.

[7] Y. Han, S. Li, Y. Cao, Q. Ma, and M. Yoshikawa, "Voice-indistinguishability: Protecting voiceprint in privacy-preserving speech data release," in *2020 IEEE International Conference on Multimedia and Expo (ICME)*. IEEE, 2020, pp. 1–6.

[8] L. Fan, "Practical image obfuscation with provable privacy," in *2019 IEEE international conference on multimedia and expo (ICME)*. IEEE, 2019, pp. 784–789.

[9] J.-W. Chen, L.-J. Chen, C.-M. Yu, and C.-S. Lu, "Perceptual indistinguishability-net (pi-net): Facial image obfuscation with manipulable semantics," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 6478–6487.

[10] Y. Zhao and J. Chen, "A survey on differential privacy for unstructured data content," *ACM Computing Surveys (CSUR)*, vol. 54, no. 10s, pp. 1–28, 2022.

[11] Y. Zhao, J. T. Du, and J. Chen, "Scenario-based adaptations of differential privacy: A technical survey," *ACM Computing Surveys*, vol. 56, no. 8, pp. 1–39, 2024.

[12] D. Desfontaines and B. Péjó, "SoK: Differential Privacies," *Proceedings on Privacy Enhancing Technologies (PoPETs)*, vol. 2020, no. 2, pp. 288–313, 2020.

[13] M. C. Tschantz, S. Sen, and A. Datta, "SoK: Differential Privacy as a Causal Property," in *2020 IEEE Symposium on Security and Privacy (SP)*. IEEE, 2020, pp. 354–371.

[14] K. Chatzikokolakis, E. ElSalamouny, C. Palamidessi, and A. Pazii, "Methods for Location Privacy: A Comparative Overview," *Foundations and Trends in Privacy and Security*, vol. 1, no. 4, pp. 199–257, 2017.

[15] M. Du, X. Yue, S. S. M. Chow, and H. Sun, "Sanitizing sentence embeddings (and labels) for local differential privacy," *Proceedings of the ACM Web Conference (WWW)*, pp. 2349–2359, 2023.

[16] C. Dwork, F. McSherry, K. Nissim, and A. Smith, "Calibrating noise to sensitivity in private data analysis," in *Theory of Cryptography: Third Theory of Cryptography Conference, TCC 2006, New York, NY, USA, March 4-7, 2006. Proceedings 3*. Springer, 2006, pp. 265–284.

[17] S. Meisenbacher, C. J. Lee, and F. Matthes, "Spend Your Budget Wisely: Towards an Intelligent Distribution of the Privacy Budget in Differentially Private Text Rewriting," in *Proceedings of the Fifteenth ACM Conference on Data and Application Security and Privacy*, pp. 84–95, 2024.

[18] R. Carpentier, B. Z. H. Zhao, H. J. Asghar, and D. Kaafar, "Preempting text sanitization utility in resource-constrained privacy-preserving LLM interactions," *arXiv preprint arXiv:2411.11521*, 2024.

[19] F. Koufogiannis, S. Han, and G. J. Pappas, "Optimality of the laplace mechanism in differential privacy," *arXiv preprint arXiv:1504.00065*, 2015.

[20] O. Feyisetan, B. Balle, T. Drake, and T. Diethe, "Privacy-and utility-preserving textual analysis via calibrated multivariate perturbations," in *Proceedings of the 13th international conference on web search and data mining*, 2020, pp. 178–186.

[21] P. Dharangutte, J. Gao, R. Gong, and F.-Y. Yu, "Integer subspace differential privacy," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 37, no. 6, 2023, pp. 7349–7357.

[22] S. P. Kasiviswanathan, H. K. Lee, K. Nissim, S. Raskhodnikova, and A. D. Smith, "What can we learn privately?" *CoRR*, vol. abs/0803.0924, 2008. [Online]. Available: http://arxiv.org/abs/0803.0924

[23] K. Chatzikokolakis, C. Palamidessi, and M. Stronati, "A predictive differentially-private mechanism for mobility traces," in *Privacy Enhancing Technologies: 14th International Symposium, PETS 2014, Amsterdam, The Netherlands, July 16-18, 2014. Proceedings 14*. Springer, 2014, pp. 21–41.

[24] A. R. Chowdhury, D. Glukhov, D. Anshumaan, P. Chalasani, N. Papernot, S. Jha, and M. Bellare, "Pr$\epsilon\epsilon$mpt: Sanitizing sensitive prompts for LLMs," *arXiv preprint arXiv:2504.05147*, 2025.

[25] K. Fawaz and K. G. Shin, "Location privacy protection for smartphone users," in *Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security*, 2014, pp. 239–250.

[26] O. Feyisetan, T. Diethe, and T. Drake, "Leveraging hierarchical representations for preserving privacy and utility in text," in *2019 IEEE International Conference on Data Mining (ICDM)*. IEEE, 2019, pp. 210–219.

[27] Y. Wang, Z. Huang, S. Mitra, and G. E. Dullerud, "Entropy-minimizing mechanism for differential privacy of discrete-time linear feedback systems," in *53rd IEEE conference on decision and control*. IEEE, 2014, pp. 2130–2135.

[28] N. Fernandes, M. Dras, and A. McIver, "Author obfuscation using generalised differential privacy," *arXiv preprint arXiv:1805.08866*, 2018.

[29] N. Fernandes, M. Dras, and A. McIver, "Generalised differential privacy for text document processing," in *Proceedings of the 8th International Conference on Principles of Security and Trust (POST 2019), ETAPS 2019*, pp. 123–148, Springer, 2019.

[30] C. Borgs, J. Chayes, A. Smith, and I. Zadik, "Revealing network structure, confidentially: Improved rates for node-private graphon estimation," in *2018 IEEE 59th Annual Symposium on Foundations of Computer Science (FOCS)*. IEEE, 2018, pp. 533–543.

[31] S. Haney, A. Machanavajjhala, and B. Ding, "Design of policy-aware differentially private algorithms," *arXiv preprint arXiv:1404.3722*, 2014.

[32] C. Qiu, A. Squicciarini, C. Pang, N. Wang, and B. Wu, "Location privacy protection in vehicle-based spatial crowdsourcing via geo-indistinguishability," *IEEE Transactions on Mobile Computing*, vol. 21, no. 7, pp. 2436–2450, 2020.

[33] Y. Cao, Y. Xiao, S. Takagi, L. Xiong, M. Yoshikawa, Y. Shen, J. Liu, H. Jin, and X. Xu, "Pglp: Customizable and rigorous location privacy through policy graph," in *European Symposium on Research in Computer Security*. Springer, 2020, pp. 655–676.

[34] B. Ma, X. Wang, W. Ni, and R. P. Liu, "Personalized location privacy with road network-indistinguishability," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 11, pp. 20 860–20 872, 2022.

[35] Y. Kawamoto and T. Murakami, "Local obfuscation mechanisms for hiding probability distributions," in *European Symposium on Research in Computer Security*. Springer, 2019, pp. 128–148.

[36] Z. Xu, A. Aggarwal, O. Feyisetan, and N. Teissier, "A differentially private text perturbation method using a regularized mahalanobis metric," *arXiv preprint arXiv:2010.11947*, 2020.

[37] Y. Zhao, D. Yuan, J. T. Du, and J. Chen, "Geo-ellipse-indistinguishability: Community-aware location privacy protection for directional distribution," *IEEE Transactions on Knowledge and Data Engineering*, vol. 35, no. 7, pp. 6957–6967, 2022.

[38] B. Weggenmann and F. Kerschbaum, "Differential privacy for directional data," in *Proceedings of the 2021 ACM SIGSAC Conference on Computer and Communications Security*, 2021, pp. 1205–1222.

[39] N. Fernandes, Y. Kawamoto, and T. Murakami, "Locality sensitive hashing with extended differential privacy," in *European Symposium on Research in Computer Security*. Springer, 2021, pp. 563–583.

[40] F. Boninsegna *et al.*, "Locality sensitive hashing of trajectories under local differential privacy." in *SEBD*, 2023, pp. 681–687.

[41] A. Brauer, V. Mäkinen, L. Ruotsalainen, and J. Oksanen, "Time will not tell: Temporal approaches for privacy-preserving trajectory publishing," *Computers, Environment and Urban Systems*, vol. 112, p. 102154, 2024.

[42] M. Yang, T. Guo, T. Zhu, I. Tjuawinata, J. Zhao, and K.-Y. Lam, "Local differential privacy and its applications: A comprehensive survey," *Computer Standards & Interfaces*, vol. 89, p. 103827, 2024.

[43] F. McSherry and K. Talwar, "Mechanism design via differential privacy," in *48th Annual IEEE Symposium on Foundations of Computer Science (FOCS'07)*. IEEE, 2007, pp. 94–103.

[44] J. Hua et al., "A geo-indistinguishable location perturbation mechanism for location-based services supporting frequent queries," *TIFS*, 2017.

[45] X. Ma et al., "Agent: An adaptive geo-indistinguishable mechanism for continuous location-based service," *Peer-to-Peer Netw. Appl.*, 2018.

[46] A. Athanasiou, K. Chatzikokolakis, and C. Palamidessi, "Enhancing metric privacy with a shuffler," in *Proceedings of the 25th Privacy Enhancing Technologies Symposium (PETS 2025)*, 2025.

[47] A. Cheu, A. Smith, J. Ullman, D. Zeber, and M. Zhilyaev, "Distributed Differential Privacy via Shuffling," in *Advances in Cryptology – EUROCRYPT 2019*. Springer, 2019, pp. 375–403.

[48] Ú. Erlingsson, V. Feldman, I. Mironov, A. Raghunathan, K. Talwar, and A. Thakurta, "Amplification by Shuffling: From Local to Central Differential Privacy via Anonymity," in *Proceedings of the Thirtieth Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*. SIAM, 2019, pp. 2468–2479.

[49] U. I. Atmaca, S. Biswas, C. Maple, and C. Palamidessi, "A Privacy-Preserving Querying Mechanism with High Utility for Electric Vehicles," *IEEE Open Journal of Vehicular Technology*, vol. 5, pp. 262–277, 2024.

[50] S. Biswas, C. Palamidessi, "PRIVIC: A Privacy-Preserving Method for Incremental Collection of Location Data," *Proceedings on Privacy Enhancing Technologies (PoPETs)*, vol. 2024, no. 1, pp. 582–602, 2024.

[51] P. Faustini, N. Fernandes, A. McIver, and C. Palamidessi, "Empirical Calibration and Metric Differential Privacy in Language Models," *IEEE Access*, vol. 13, 2025.

[52] D. Agrawal and C. C. Aggarwal, "On the Design and Quantification of Privacy Preserving Data Mining Algorithms," in *Proceedings of the 20th ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems (PODS)*, 2001, pp. 247–255.

[53] P. Kairouz, S. Oh, and P. Viswanath, "Extremal Mechanisms for Local Differential Privacy," in *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 27, 2014, pp. 2879–2887.

[54] T. Wang, J. Blocki, N. Li, and S. Jha, "Locally Differentially Private Protocols for Frequency Estimation," in *Proceedings of the 26th USENIX Security Symposium*, 2017, pp. 729–745.

[55] E. ElSalamouny and C. Palamidessi, "Full Convergence of the Iterative Bayesian Update and Applications to Mechanisms for Privacy Protection," in *Proceedings of the 5th IEEE European Symposium on Security and Privacy (EuroS&P)*, 2020, pp. 490–507.

[56] H. H. Arcolezi, S. Cerna, and C. Palamidessi, "On the Utility Gain of Iterative Bayesian Update for Locally Differentially Private Mechanisms," in *Data and Applications Security and Privacy XXXVII (DBSec)*, ser. Lecture Notes in Computer Science, vol. 13942. Springer, 2023, pp. 165–183.

[57] E. ElSalamouny and C. Palamidessi, "On the Consistency and Performance of the Iterative Bayesian Update," *arXiv preprint arXiv:2508.09980*, 2025.

[58] J. Imola, A. Roy Chowdhury, and K. Chaudhuri, "Metric Differential Privacy at the User-Level via the Earth Mover's Distance," in *Proceedings of the 2024 ACM SIGSAC Conference on Computer and Communications Security (CCS)*, 2024.

[59] R. Chen, L. Li, Y. Ma, Y. Gong, Y. Guo, T. Ohtsuki, and M. Pan, "Constructing mobile crowdsourced covid-19 vulnerability map with geo-indistinguishability," *IEEE Internet of Things Journal*, vol. 9, no. 18, pp. 17 403–17 416, 2022.

[60] R. Mendes and J. Vilela, "On the effect of update frequency on geo-indistinguishability of mobility traces," in *Proceedings of the 11th ACM Conference on Security & Privacy in Wireless and Mobile Networks*, 2018, pp. 271–276.

[61] A. M. Awon, Y. Lu, S. Potka, and A. Thomo, "CluSanT: Differentially private and semantically coherent text sanitization," *Proceedings of the 2025 Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT)*, pp. 3676–3693, 2025.

[62] H. To, C. Shahabi, and L. Xiong, "Privacy-preserving online task assignment in spatial crowdsourcing with untrusted server," in *2018 IEEE 34th international conference on data engineering (ICDE)*. IEEE, 2018, pp. 833–844.

[63] C. Yang, X. Ju, E. Liu, Y. Geng, and R. Wang, "Blockchain-based indoor location paging and answering service with truncated-geo-indistinguishability," *IET Blockchain*, vol. 1, no. 2-4, pp. 105–117, 2021.

[64] P. Huang, X. Zhang, L. Guo, and M. Li, "Incentivizing crowdsensing-based noise monitoring with differentially-private locations," *IEEE Transactions on Mobile Computing*, vol. 20, no. 2, pp. 519–532, 2019.

[65] F. Li, J. Dong, M. Chen, and P. Li, "A privacy preserving method for trajectory data publishing based on geo-indistinguishability," in *International Conference on Advanced Data Mining and Applications*. Springer, 2023, pp. 633–647.

[66] M. Li, Y. Zeng, L. Zheng, L. Chen, and Q. Li, "Accurate and efficient trajectory-based contact tracing with secure computation and geo-indistinguishability," in *International Conference on Database Systems for Advanced Applications*. Springer, 2023, pp. 300–316.

[67] G. Duarte et al., "A privacy-aware remapping mechanism for location data," in *SAC*, 2024.

[68] M. Min, H. Zhu, J. Ding, S. Li, L. Xiao, M. Pan, and Z. Han, "Personalized 3d location privacy protection with differential and distortion geo-perturbation," *IEEE Transactions on Dependable and Secure Computing*, vol. 21, no. 4, pp. 3629–3643, 2023.

[69] L. Yu, S. Zhang, Y. Meng, S. Du, Y. Chen, Y. Ren, and H. Zhu, "Privacy-preserving location-based advertising via longitudinal geo-indistinguishability," *IEEE Transactions on Mobile Computing*, vol. 23, no. 8, pp. 8256–8273, 2023.

[70] R. Al-Dhubhani and J. M. Cazalas, "An adaptive geo-indistinguishability mechanism for continuous lbs queries," *Wireless Networks*, vol. 24, no. 8, pp. 3221–3239, 2018.

[71] M. Cunha, R. Mendes, and J. P. Vilela, "Clustering geo-indistinguishability for privacy of continuous location traces," in *2019 4th International Conference on Computing, Communications and Security (ICCCS)*. IEEE, 2019, pp. 1–8.

[72] N. Fernandes, A. McIver, and C. Morgan, "The laplace mechanism has optimal utility for differential privacy over continuous queries," in *2021 36th Annual ACM/IEEE Symposium on Logic in Computer Science (LICS)*. IEEE, 2021, pp. 1–12.

[73] S. Oya, C. Troncoso, and F. Pérez-González, "Is geo-indistinguishability what you are looking for?" in *Proc. of the 2017 on Workshop on Privacy in the Electronic Society*, ser. WPES '17. New York, NY, USA: Association for Computing Machinery, 2017, p. 137–140. [Online]. Available: https://doi.org/10.1145/3139550.3139555

[74] K. Chatzikokolakis, E. Elsalamouny, and C. Palamidessi, "Efficient utility improvement for location privacy," *Proceedings on Privacy Enhancing Technologies*, 2017.

[75] A. Pankova and P. Laud, "Interpreting epsilon of differential privacy in terms of advantage in guessing or approximating sensitive attributes," in *2022 IEEE 35th Computer Security Foundations Symposium (CSF)*. IEEE, 2022, pp. 96–111.

[76] F. Koufogiannis, S. Han, and G. J. Pappas, "Gradual release of sensitive data under differential privacy," *Journal of Privacy and Confidentiality*, vol. 7, no. 2, 2016.

[77] F. Koufogiannis and G. J. Pappas, "Location-dependent privacy," in *2016 IEEE 55th Conference on Decision and Control (CDC)*. IEEE, 2016, pp. 7586–7591.

[78] R. Ahuja, G. Ghinita, and C. Shahabi, "A utility-preserving and scalable technique for protecting location data with geo-indistinguishability," in *22nd International Conference on Extending Database Technology, EDBT 2019*. OpenProceedings.org, 2019, pp. 217–228.

[79] M. Min, H. Zhu, S. Li, H. Zhang, L. Xiao, M. Pan, and Z. Han, "Semantic adaptive geo-indistinguishability for location privacy protection in mobile networks," *IEEE Transactions on Vehicular Technology*, 2024.

[80] M. Min, L. Xiao, J. Ding, H. Zhang, S. Li, M. Pan, and Z. Han, "3d geo-indistinguishability for indoor location-based services," *IEEE Transactions on Wireless Communications*, vol. 21, no. 7, pp. 4682–4694, 2021.

[81] A. Fathalizadeh, V. Moghtadaiee, and M. Alishahi, "Indoor geo-indistinguishability: Adopting differential privacy for indoor location data protection," *IEEE Transactions on Emerging Topics in Computing*, vol. 12, no. 1, pp. 293–306, 2023.

[82] W. Eltarjaman, R. Dewri, and R. Thurimella, "Location privacy for rank-based geo-query systems," *Proceedings on Privacy Enhancing Technologies*, 2017.

[83] P. Kamalaruban, V. Perrier, H. J. Asghar, and M. A. Kaafar, "Not all attributes are created equal: dx-private mechanisms for linear queries," *Proceedings on Privacy Enhancing Technologies*, 2020.

[84] S. Arnold, D. Yesilbas, and S. Weinzierl, "Guiding text-to-text privatization by syntax," in *Proceedings of the 3rd Workshop on Trustworthy Natural Language Processing (TrustNLP 2023)*, 2023, pp. 151–162.

[85] F. Fioretto, T. W. Mak, and P. Van Hentenryck, "Privacy-preserving obfuscation of critical infrastructure networks," *arXiv preprint arXiv:1905.09778*, 2019.

[86] L. Fan and L. Bonomi, "Time series sanitization with metric-based privacy," in *2018 IEEE International Congress on Big Data (BigData Congress)*. IEEE, 2018, pp. 264–267.

[87] W. Tong, J. Hua, and S. Zhong, "A jointly differentially private scheduling protocol for ridesharing services," *IEEE Transactions on Information Forensics and Security*, vol. 12, no. 10, pp. 2444–2456, 2017.

[88] C. Huang, R. Lu, H. Zhu, J. Shao, A. Alamer, and X. Lin, "Eppd: Efficient and privacy-preserving proximity testing with differential privacy techniques," in *2016 IEEE International Conference on Communications (ICC)*. IEEE, 2016, pp. 1–6.

[89] X. Shi, D. Yu, M. Fu, and W.-A. Zhang, "Clap: A contract-based incentive mechanism for cooperative localization balancing localization accuracy and location privacy," *IEEE Internet of Things Journal*, vol. 9, no. 9, pp. 6678–6687, 2021.

[90] F. Galli, S. Biswas, K. Jung, T. Cucinotta, C. Palamidessi *et al.*, "Group privacy for personalized federated learning," *ICISSP*, vol. 1, pp. 252–263, 2023.

[91] F. Galli et al., "Advancing personalized federated learning: Group privacy, fairness, and beyond," *SN Computer Science*, 2023.

[92] K. Chatzikokolakis, C. Palamidessi, and M. Stronati, "Constructing elastic distinguishability metrics for location privacy," *Proceedings on Privacy Enhancing Technologies*, 2015.

[93] S. Takagi, Y. Cao, Y. Asano, and M. Yoshikawa, "Geo-graph-indistinguishability: Location privacy on road networks with differential privacy," *IEICE TRANSACTIONS on Information and Systems*, vol. 106, no. 5, pp. 877–894, 2023.

[94] B. Ma, X. Lin, X. Wang, B. Liu, Y. He, W. Ni, and R. P. Liu, "New cloaking region obfuscation for road network-indistinguishability and location privacy," in *Proceedings of the 25th International Symposium on Research in Attacks, Intrusions and Defenses*, 2022, pp. 160–170.

[95] L. Yu, L. Liu, and C. Pu, "Dynamic differential location privacy with personalized error bounds." in *NDSS*, vol. 17, 2017, pp. 1–15.

[96] Z. Shun, D. Benfei, C. Zhili, and Z. Hong, "On the differential privacy of dynamic location obfuscation with personalized error bounds," *arXiv preprint arXiv:2101.12602*, 2021.

[97] S. Zhang, P. Lan, B. Duan, Z. Chen, H. Zhong, and N. N. Xiong, "Dpive: a regionalized location obfuscation scheme with personalized privacy levels," *ACM Transactions on Sensor Networks*, vol. 20, no. 2, pp. 1–26, 2024.

[98] M. E. Gursoy, A. Tamersoy, S. Truex, W. Wei, and L. Liu, "Secure and utility-aware data collection with condensed local differential privacy," *IEEE Transactions on Dependable and Secure Computing*, vol. 18, no. 5, pp. 2365–2378, 2019.

[99] Y. Ren, X. Li, Y. Miao, R. H. Deng, J. Weng, S. Ma, and J. Ma, "Distpreserv: Maintaining user distribution for privacy-preserving location-based services,"

IEEE Transactions on Mobile Computing, vol. 22, no. 6, pp. 3287–3302, 2022.
[100] X. Dong et al., "Preserving geo-indistinguishability of the primary user in dynamic spectrum sharing," *IEEE Trans. on Vehicular Technology*, 2019.
[101] B. Niu, Y. Chen, Z. Wang, F. Li, B. Wang, and H. Li, "Eclipse: Preserving differential location privacy against long-term observation attacks," *IEEE Transactions on Mobile Computing*, vol. 21, no. 1, pp. 125–138, 2020.
[102] X. Li, Y. Ren, L. T. Yang, N. Zhang, B. Luo, J. Weng, and X. Liu, "Perturbation-hidden: Enhancement of vehicular privacy for location-based services in internet of vehicles," *IEEE Transactions on Network Science and Engineering*, vol. 8, no. 3, pp. 2073–2086, 2020.
[103] S. Wang, Y. Nie, P. Wang, H. Xu, W. Yang, and L. Huang, "Local private ordinal data distribution estimation," in *IEEE INFOCOM 2017-IEEE Conference on Computer Communications*. IEEE, 2017, pp. 1–9.
[104] A. M. Awon, Y. Lu, S. Potka, and A. Thomo, "CluSanT: Differentially Private and Semantically Coherent Text Sanitization," in *Proceedings of the 2025 Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)*, pp. 3676–3693, 2025.
[105] X. Yue, M. Du, T. Wang, Y. Li, H. Sun, and S. S. Chow, "Differential privacy for text analytics via natural text sanitization," in *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, 2021, pp. 3853–3866.
[106] O. Feyisetan, A. Aggarwal, Z. Xu, and N. Teissier, "Research challenges in designing differentially private text generation mechanisms," *arXiv preprint arXiv:2012.05403*, 2020.
[107] S. Raskhodnikova and A. Smith, "Lipschitz extensions for node-private graph statistics and the generalized exponential mechanism," in *2016 IEEE 57th Annual Symposium on Foundations of Computer Science (FOCS)*. IEEE, 2016, pp. 495–504.
[108] A. Roy Chowdhury, B. Ding, S. Jha, W. Liu, and J. Zhou, "Strengthening order preserving encryption with differential privacy," in *Proceedings of the 2022 ACM SIGSAC Conference on Computer and Communications Security*, 2022, pp. 2519–2533.
[109] S. Zhang, T. Zhang, Z. Chen, and N. Xiong, "Geo-moea: A multi-objective evolutionary algorithm with geo-obfuscation for mobile crowdsourcing workers," *arXiv preprint arXiv:2201.11300*, 2022.
[110] N. E. Bordenabe, K. Chatzikokolakis, and C. Palamidessi, "Optimal geo-indistinguishable mechanisms for location privacy," in *Proceedings of the 2014 ACM SIGSAC conference on computer and communications security*, 2014, pp. 251–262.
[111] C. Qiu, "Enhancing scalability of metric differential privacy via secret dataset partitioning and benders decomposition," in *Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence*, 2024, pp. 1944–1952.
[112] C. Qiu, "Interpolation-Based Optimization for Enforcing ℓ_p-Norm Metric Differential Privacy in Continuous and Fine-Grained Domains," in *Proceedings of the 35th USENIX Security Symposium*, 2026.
[113] Y. Lin, "Geo-indistinguishable masking: enhancing privacy protection in spatial point mapping," *Cartography and Geographic Information Science*, vol. 50, no. 6, pp. 608–623, 2023.
[114] P. Pappachan, C. Qiu, A. Squicciarini, and V. S. H. Manjunath, "User customizable and robust geo-indistinguishability for location privacy," in *EDBT*, 2023.
[115] L. Wang, D. Yang, X. Han, T. Wang, D. Zhang, and X. Ma, "Location privacy-preserving task allocation for mobile crowdsensing with differential geo-obfuscation," in *Proceedings of the 26th International Conference on World Wide Web*, 2017, pp. 627–636.
[116] L. Wang, D. Yang, X. Han, D. Zhang, and X. Ma, "Mobile crowdsourcing task allocation with differential-and-distortion geo-obfuscation," *IEEE Transactions on Dependable and Secure Computing*, vol. 18, no. 2, pp. 967–981, 2019.
[117] M. Abadi, A. Chu, I. Goodfellow, H. B. McMahan, I. Mironov, K. Talwar, and L. Zhang, "Deep learning with differential privacy," in *Proc. ACM SIGSAC Conf. Computer and Communications Security (CCS)*, pp. 308–318, 2016.
[118] R. Shokri, G. Theodorakopoulos, J.-Y. Le Boudec, and J.-P. Hubaux, "Quantifying location privacy," in *Proc. IEEE Symp. Security and Privacy (S&P)*, pp. 247–262, 2011.
[119] H. Xiao and S. Devadas, "PAC privacy: automatic privacy measurement and control of data processing," *arXiv preprint arXiv:2210.03458*, 2023.
[120] Y. Qian, Y. Ma, J. Chen, D. Wu, D. Tian, and K. Hwang, "Optimal location privacy preserving and service quality guaranteed task allocation in vehicle-based crowdsensing networks," *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no. 7, pp. 4367–4375, 2021.
[121] C. Qiu, A. Squicciarini, Z. Li, C. Pang, and L. Yan, "Time-efficient geo-obfuscation to protect worker location privacy over road networks in spatial crowdsourcing," in *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*, 2020, pp. 1275–1284.
[122] P. Zhang, X. Cheng, S. Su, and N. Wang, "Task allocation under geo-indistinguishability via group-based noise addition," *IEEE Transactions on Big Data*, vol. 9, no. 3, pp. 860–877, 2022.
[123] C. Qiu, S. Yadav, Y. Ji, A. C. Squicciarini, R. Dantu, J. Zhao, and C.-Z. Xu, "Fine-grained geo-obfuscation to protect workers' location privacy in time-sensitive spatial crowdsourcing." in *EDBT*, 2024, pp. 373–385.
[124] P. Zhang, C. Hu, D. Chen, H. Li, and Q. Li, "Shiftroute: Achieving location privacy for map services on smartphones," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 5, pp. 4527–4538, 2018.
[125] D. Yu, X. Shi, L. Chai, W.-A. Zhang, and J. Chen, "Balancing localization accuracy and location privacy in mobile cooperative localization," *IEEE Transactions on Signal Processing*, vol. 71, pp. 2804–2818, 2023.
[126] M. Min, H. Zhu, S. Yang, J. Xu, J. Tong, S. Li, and J. Shu, "Geo-perturbation for task allocation in 3-d mobile crowdsourcing: An a3c-based approach," *IEEE Internet of Things Journal*, vol. 11, no. 2, pp. 1854–1865, 2023.
[127] J. Liu, C. Zhang, and Y. Fang, "Epic: A differential privacy framework to defend smart homes against internet traffic analysis," *IEEE Internet of Things Journal*, vol. 5, no. 2, pp. 1206–1217, 2018.
[128] C. Qiu, L. Yan, A. Squicciarini, J. Zhao, C. Xu, and P. Pappachan, "Trafficadaptor: an adaptive obfuscation strategy for vehicle location privacy against traffic flow aware attacks," in *Proceedings of the 30th International Conference on Advances in Geographic Information Systems*, 2022, pp. 1–10.
[129] M. Boedihardjo, T. Strohmer, and R. Vershynin, "Metric geometry of the privacy-utility tradeoff," *arXiv preprint arXiv:2405.00329*, 2024.
[130] M. A. Hoefler, K. Mueller, and W. Samek, "FedXDS: Leveraging Model Attribution Methods to counteract Data Heterogeneity in Federated Learning," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 4572–4581, 2025.
[131] R. Shokri, G. Theodorakopoulos, C. Troncoso, J. Hubaux, and J. L. Boudec, "Protecting location privacy: Optimal strategy against localization attacks," in *Proc. of ACM CCS*, 2012, pp. 617–627.
[132] R. Shokri, "Privacy games: Optimal user-centric data obfuscation," *PETS*, 2015.
[133] L. Wang, D. Zhang, D. Yang, B. Y. Lim, and X. Ma, "Differential location privacy for sparse mobile crowdsensing," in *2016 IEEE 16th International Conference on Data Mining (ICDM)*, 2016, pp. 1257–1262.
[134] L. Wang, D. Yang, X. Han, T. Wang, D. Zhang, and X. Ma, "Location privacy-preserving task allocation for mobile crowdsensing with differential geo-obfuscation," in *Proc. of ACM WWW*, 2017, pp. 627–636.
[135] S. Oya, C. Troncoso, and F. Pérez-González, "Back to the drawing board: Revisiting the design of optimal location privacy-preserving mechanisms," in *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*, 2017, pp. 1959–1972.
[136] M. S. Alvim et al., "Metric-based local differential privacy for statistical applications," *arXiv*, 2018.
[137] R. Mendes, M. Cunha, and J. Vilela, "Impact of frequency of location reports on the privacy level of geo-indistinguishability," *Proceedings on Privacy Enhancing Technologies*, vol. 2020, pp. 379–396, 04 2020.
[138] Y. Da, R. Ahuja, L. Xiong, and C. Shahabi, "React: real-time contact tracing and risk monitoring via privacy-enhanced mobile tracking," in *2021 IEEE 37th International Conference on Data Engineering (ICDE)*. IEEE, 2021, pp. 2729–2732.
[139] E. Wilson, V. Bindschaedler, S. Jörg, S. Sheikholeslam, K. Butler, and E. Jain, "Towards Privacy-preserving Photorealistic Self-avatars in Mixed Reality," *arXiv preprint arXiv:2507.22153*, 2025.
[140] C. Qiu, R. Liu, P. Pappachan, A. Squicciarini, and X. Xie, "Time-efficient locally relevant geo-location privacy protection," *Proceedings on Privacy Enhancing Technologies*, 2025.
[141] R. Liu and C. Qiu, "Panda: Rethinking metric differential privacy optimization at scale with anchor-based approximation," in *Proceedings of The 32nd ACM Conference on Computer and Communications Security (CCS)*, 2025.
[142] R. S. Carvalho, T. Vasiloudis, O. Feyisetan, and K. Wang, "Tem: High utility metric differential privacy on text," in *Proceedings of the 2023 SIAM International Conference on Data Mining (SDM)*. SIAM, 2023, pp. 883–890.
[143] S. Takagi, Y. Cao, Y. Asano, and M. Yoshikawa, "Geo-graph-indistinguishability: Protecting location privacy for lbs over road networks," in *Data and Applications Security and Privacy XXXIII: 33rd Annual IFIP WG 11.3 Conference, DBSec 2019, Charleston, SC, USA, July 15–17, 2019, Proceedings 33*. Springer, 2019, pp. 143–163.
[144] P. Zhang et al., "Area coverage-based worker recruitment under geo-indistinguishability," *Computer Networks*, 2022.
[145] S. Meisenbacher et al., "Efficient and utility-preserving text obfuscation leveraging word-level metric differential privacy," *arXiv*, 2024.
[146] H. Yan, X. Li, W. Zhang, Q. Chen, B. Wang, H. Li, and X. Lin, "Coder: Protecting privacy in image retrieval with differential privacy," *IEEE Transactions on Dependable and Secure Computing*, 2024.
[147] Y. Zheng, L. Zhang, X. Xie, and W.-Y. Ma, "Mining interesting locations and travel sequences from gps trajectories," in *Proceedings of the 19th international conference on World wide web*. ACM, 2010, pp. 791–800.
[148] R. Mendes, M. Cunha, and J. P. Vilela, "Impact of frequency of location reports on the privacy level of geo-indistinguishability," *Proceedings on Privacy Enhancing Technologies*, 2020.
[149] J. Yuan, Y. Zheng, C. Zhang, W. Xie, X. Xie, G. Sun, and Y. Huang, "T-drive: driving directions based on taxi trajectories," in *Proceedings of the 18th SIGSPATIAL International conference on advances in geographic information systems*, 2010, pp.

99–108.
[150] E. Cho, S. A. Myers, and J. Leskovec, "Friendship and mobility: User movement in location-based social networks," in *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2011, pp. 1082–1090.
[151] S. Oya, C. Troncoso, and F. Pérez-González, "Is geo-indistinguishability what you are looking for?" in *Proceedings of the 2017 on Workshop on Privacy in the Electronic Society*, 2017, pp. 137–140.
[152] D. Yang, D. Zhang, V. W. Liu, and H. Zha, "A categorial and geographical-aware recommender system," in *Proceedings of the 22nd International Conference on World Wide Web*. ACM, 2013, pp. 1447–1458.
[153] Z. Wang, J. Hu, R. Lv, J. Wei, Q. Wang, D. Yang, and H. Qi, "Personalized privacy-preserving task allocation for mobile crowdsensing," *IEEE Transactions on Mobile Computing*, vol. 18, no. 6, pp. 1330–1341, 2018.
[154] M. Piorkowski, N. Sarafijanovic-Djukic, and M. Grossglauser, "Hail cabs in the air: A large-scale study of gps-enabled taxis," in *Proceedings of the 9th ACM SIGCOMM conference on Internet measurement conference*, 2009, pp. 386–398.
[155] L. Moreira-Matias, J. Gama, M. Ferreira, J. Mendes-Moreira, and L. Damas, "Predicting taxi-passenger demand using streaming data," in *IEEE Transactions on Intelligent Transportation Systems*, vol. 14, no. 3. IEEE, 2013, pp. 1393–1402.
[156] F. M. Harper and J. A. Konstan, "The movielens datasets: History and context," *Acm transactions on interactive intelligent systems (tiis)*, vol. 5, no. 4, pp. 1–19, 2015.
[157] Yelp Inc., "Yelp Dataset," https://www.yelp.com/dataset, 2024, accessed: [Your Access Date].
[158] G. Cohen, S. Afshar, J. Tapson, and A. Van Schaik, "Emnist: Extending mnist to handwritten letters," in *2017 international joint conference on neural networks (IJCNN)*. IEEE, 2017, pp. 2921–2926.
[159] A. L. Maas, R. E. Daly, P. T. Pham, D. Huang, A. Y. Ng, and C. Potts, "Learning word vectors for sentiment analysis," in *Proceedings of the 49th annual meeting of the Association for Computational Linguistics: Human Language Technologies*, 2011, pp. 142–150.
[160] M. Potthast, B. Stein, A. Barrón-Cedeño, and P. Rosso, "Overview of the 3rd international competition on plagiarism detection," in *Notebook for PAN at CLEF 2011*, 2011.
[161] E. Stamatatos, M. Kestemont, B. Stein, and M. Potthast, "Overview of the pan 2012 traditional authorship attribution task," in *CLEF (Online Working Notes/Labs/Workshop)*, 2012.
[162] B. Pang and L. Lee, "Seeing stars: Exploiting class relationships for sentiment categorization with respect to rating scales," in *Proceedings of the 43rd annual meeting of the Association for Computational Linguistics (ACL'05)*, 2005, pp. 115–124.
[163] M. Hu and B. Liu, "Mining and summarizing customer reviews," in *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, 2004, pp. 168–177.
[164] SimpleMaps, "U.S. Cities Database," https://simplemaps.com/data/us-cities, 2023, a commonly used commercial/free database for U.S. city locations and populations.
[165] U.S. Census Bureau, "American Community Survey (ACS)," https://www.census.gov/programs-surveys/acs, 2023, ongoing survey, year of data used should be specified.
[166] Twitter, Inc. (now X Corp.), "Twitter API," https://developer.twitter.com/, 2023, data is collected via the official API; specific datasets are usually not public.
[167] Z. Yang, W. W. Cohen, and R. Salakhutdinov, "Revisiting semi-supervised learning with graph embeddings," in *International conference on machine learning*. PMLR, 2016, pp. 40–48, this paper popularized the use of citation network datasets like Cora, CiteSeer, and DBLP for graph-based learning tasks.
[168] M. Min, S. Yang, H. Zhang, J. Ding, G. Peng, M. Pan, and Z. Han, "Indoor semantic location privacy protection with safe reinforcement learning," *IEEE Transactions on Cognitive Communications and Networking*, vol. 9, no. 5, pp. 1385–1398, 2023.
[169] OpenStreetMap contributors, "OpenStreetMap," https://www.openstreetmap.org, 2024, data licensed under the Open Data Commons Open Database License (ODbL).
[170] D. L. Balk, F. Pozzi, G. Yetman, U. Deichmann, and A. Nelson, "GRUMPv1: Global Rural-Urban Mapping Project, Version 1," Columbia University. Palisades, NY, 2004, https://sedac.ciesin.columbia.edu/data/set/grump-v1-population-density.
[171] Google, "Google Places API," https://developers.google.com/maps/documentation/places, 2024, part of the Google Maps Platform.
[172] E. ElSalamouny, K. Chatzikokolakis, and C. Palamidessi, "Generalized differential privacy: Regions of priors that admit robust optimal mechanisms," *Horizons of the Mind. A Tribute to Prakash Panangaden: Essays Dedicated to Prakash Panangaden on the Occasion of His 60th Birthday*, pp. 292–318, 2014.
[173] E. ElSalamouny et al., "Differential privacy models for location-based services," *Transactions on Data Privacy*, 2016.
[174] K. Odoh, "Group-wise k-anonymity meets (ε, δ) differentially privacy scheme," in *Companion Proceedings of the ACM Web Conference 2024*, 2024, pp. 802–805.
[175] H. J. Asghar, R. Carpentier, B. Z. H. Zhao, and D. Kaafar, "d_X-privacy for text and the curse of dimensionality," *arXiv preprint arXiv:2411.13784*, 2024.
[176] A. Roy Chowdhury *et al.*, "Pr$\epsilon\epsilon$mpt: Sanitizing sensitive prompts for LLMs," in *Proc. Network and Distributed System Security Symposium (NDSS)*, 2026.
[177] R. Harel, N. Gilboa, and Y. Pinter, "Token-level privacy in large language models," *arXiv preprint arXiv:2503.03652*, 2025.
[178] P. Mai, R. Yan, Z. Huang, Y. Yang, and Y. Pang, "Split-and-denoise: Protect large language model inference with local differential privacy," in *Proc. International Conference on Machine Learning (ICML)*, 2024, pp. 34281–34302.
[179] Z. Zeng, J. Wang, J. Yang, Z. Lu, H. Li, H. Zhuang, and C. Chen, "PrivacyRestore: Privacy-preserving inference in large language models via privacy removal and restoration," in *Proc. Annual Meeting of the Association for Computational Linguistics (ACL)*, 2025.
[180] F. Tian, P. Bhattacharjee, H. Hanson, G. D. Rubin, J. Y. Lo, and R. Tandon, "STAMP: Selective task-aware mechanism for text privacy," in *Proc. European Chapter of the Association for Computational Linguistics (EACL)*, 2026.
[181] S. Meisenbacher, A. Nagy, M. Chevli, and F. Matthes, "The double-edged sword of LLM-based data reconstruction: Understanding and mitigating contextual vulnerability in word-level differential privacy text sanitization," in *Proc. 24th Workshop on Privacy in the Electronic Society (WPES)*, 2025.
[182] S. Pang, Z. Lu, H. Wang, P. Fu, Y. Zhou, M. Xue, and B. Li, "Reconstruction of differentially private text sanitization via large language models," *arXiv preprint arXiv:2410.12443*, 2024.
[183] D. Bollegala, S. Otake, T. Machide, and K.-I. Kawarabayashi, "A metric differential privacy mechanism for sentence embeddings," *ACM Transactions on Privacy and Security*, vol. 28, no. 2, 2025.
[184] Y. Li, Z. Tan, P. Branco, and Y. Liu, "Privacy-preserving parameter-efficient fine-tuning for large language model services," *IEEE Transactions on Dependable and Secure Computing*, 2024.

A Appendix

Paper	Distance Type	Application: Description
PETS 2013 [4], PETS 2014 [23], CCS 2014 [25], UAI [3], ICDM 2019 [26]	Euclidean Distance	Location-based services (LBS) and natural language processing (NLP): $d_{\mathcal{X}}(x, x') = \sqrt{\sum_i (x_i - x'_i)^2}$ quantifies the straight-line displacement between two data points x, x', where x_i and x'_i represent the i-th coordinate of x and x', respectively
CDC 2014 [27]	Manhattan Distance	Control Systems: $d_{\mathcal{X}}(x, x') = \sum_i \lvert x_i - x'_i \rvert$ measures the total absolute difference across all coordinate dimensions.
arXiv 2018 [28], POST 2019 [29]	Word Mover's Distance (Kantorovich Distance)	NLP: $d_{\mathcal{X}}(x, x')$ measures the minimum transport cost needed to move words from one text document x to another x'. The cost is computed based on the semantic similarity of the words, measured using distances between their vector representations in a shared embedding space.
FOCS 2018 [30]	Node-Distance	Graph processing: $d_{\mathcal{X}}(x, x')$ measures the minimum number of node modifications such as additions, deletions, or substitutions needed to transform one graph structure x into another x'.
VLDB 2015 [31], TMC 2020 [32], ESORICS 2020 [33], TITS 2022 [34]	Shortest Path Distance	LBS: $d_{\mathcal{X}}(x, x')$ measures the minimum travel distance between two points x, x' along a network, considering the road or graph structure connecting them.
ESORICS 2019 [35]	Wasserstein Distance	LBS: $d_{\mathcal{X}}(x, x')$ measures the minimal cost of transforming one probability distribution x into another x', based on optimal transport theory.
arXiv 2020 [36], TKDE 2023 [37]	Regularized Mahalanobis Norm	LBS and NLP: $d_{\mathcal{X}}(x, x')$ measures distance between x and x' while considering correlations between their features, regularized to balance sensitivity and robustness in privacy mechanisms.
ICME 2020 [7], CCS 2021 [38], ESORICS 2021 [39] arXiv 2025[139]	Angular Distance	Voice processing and LBS: $d_{\mathcal{X}}(x, x')$ measures the difference in orientation between vectors x, x', often using cosine similarity to quantify differences in feature representations.
SEBD 2023 [40]	Fréchet Distance	LBS: $d_{\mathcal{X}}(x, x')$ measures the similarity between two curves x, x' by considering the minimal effort required to continuously transform one into the other while maintaining order.
arXiv 2024 [41]	Temporal Distance	LBS: $d_{\mathcal{X}}(x, x')$ measures the difference between time points or temporal sequences x and x', capturing the deviation in timestamps along trajectories.

Table 6: Metric distance definitions and descriptions in mDP works.

Mechanism Taxonomy Tables (HDM / NHDM / OPM, by Recipe DOF)

The three tables below organize HDM, NHDM, and OPM works (beyond the canonical anchor of each section) by which design degree-of-freedom of the mechanism recipe is varied. Many papers contribute to multiple cells; we list each under its primary contribution. The narrative summaries in Sec. 3.1–3.3 cite these tables and walk through their organization rather than enumerating individual works.

Table 7: HDM beyond the planar-Laplace anchor (Sec. 3.1), organized by which ingredient of the recipe ($\Pr[y|x] \propto e^{-\varepsilon\, d(x,y)}$, optionally post-processed and composed over time) is varied.

Recipe DOF	Variation & key technical idea	Representative works
(A) Metric / domain. *Plug a new d into the same Laplace recipe.*	3D Euclidean for indoor LBS; Word Mover's Distance for text/author obfuscation; per-token embedding metrics (semantic, syntactic, dimension-reduced) for NLP; GAN-latent perceptual metric for facial images; Wasserstein on probability distributions; graph node-distance for infrastructure obfuscation and graphon estimation; similarity metrics on IoT time series. Unifying point: HDM is parameterized by d.	[6, 9, 26, 28–30, 35, 80, 81, 84–86]
(B) Privacy-budget structure. *Make ε context-dependent.*	Spatially varying $\varepsilon(x)$ via the Eikonal PDE; tile-based partitions with per-tile budgets; semantic-adaptive levels driven by land use; per-attribute d_X-private mechanisms for linear queries; policy graphs (Blowfish, PGLP) enumerating which input pairs must remain indistinguishable; per-query-rank budgets; truncated/region-restricted Geo-Ind that excludes inadmissible outputs by construction; hexagon-discretized release.	[31, 33, 44, 63, 77–79, 82, 83]
(C) Sequential composition. *Cope with continuous traces; resist budget decay across time.*	The foundational predictive mechanism reusing correlated noise on predictable mobility [23]; correlation-adaptive ε between successive reports; trajectory clustering and stay-point obfuscation; per-segment perturbation of trajectory key-points; longitudinal/lifelong release; window-based adaptive allocation in 3D-spatiotemporal trajectories; COVID-era contact tracing. Shared theme: $\sum_t \varepsilon_t$ decays naively unless temporal redundancy is exploited.	[23, 59, 60, 65, 66, 68–71]
(D) Theory & post-processing. *Refine the output, characterize optimality.*	Bayesian remapping that uses a prior to relocate noisy outputs without breaking privacy; privacy-preserving variants absorbing the remap into the analysis; gradual release; entropy-view optimality for discrete-time systems; Lipschitz-privacy MSE-optimality of Laplace for ℓ_2 identity queries; continuous-domain extension to general queries; critical assessment under realistic priors; ε semantics in adversarial-advantage terms.	[19, 27, 67, 72–76]
(E) System integration. *Use HDM as the privacy substrate inside a larger system.*	Privacy-preserving task assignment in spatial crowdsourcing (workers and tasks both perturbed before matching); cooperative-localization incentives under information asymmetry; federated-learning group privacy via d-privacy that preserves topological similarity; blockchain-coordinated indoor LBS combining truncated Geo-Ind with smart-contract reward to enforce on-building outputs at the protocol level. Primary novelty is the system; HDM is the privacy primitive.	[45, 62, 64, 87–91]

Table 8: NHDM beyond the EM anchor (Sec. 3.2), organized by which ingredient of the EM recipe ($\Pr[y|x] \propto e^{\varepsilon\, u(x,y)/(2\Delta u)}$, normalized over a finite candidate set $\mathcal{Y}$, with default $u = -d$) is varied.

Recipe DOF	Variation & key technical idea	Representative works
(A′) Candidate set $\mathcal{Y}$ & its metric. *What discrete domain does EM normalize over, and what notion of similarity ranks candidates?*	Road-graph nodes with shortest-path distance (Geo-Graph-Indistinguishability / GEM); sensitivity-bipartitioned vocabulary (SanText separates sensitive from non-sensitive tokens before EM); hierarchical cluster-then-token sets (CluSanT unifies SanText and CusText via two-stage cluster→token sampling); ordinal subsets (Subset Exponential Mechanism); anisotropic ellipse-shaped output regions for directional traffic (Geo-Ellipse-Indistinguishability); free-form word-embedding spaces under metric LDP.	[20, 37, 93, 103–105, 143]
(B′) Score function u & budget allocation. *Beyond the canonical $u = -d$ and uniform ε.*	Distribution-preserving scores (DistPreserv); density-modulated scores calibrated to local embedding density; order-preserving scores strengthening Order-Preserving Encryption; linguistic-importance-weighted per-token budget allocation (*Spend Your Budget Wisely* argues uniform ε-per-token is suboptimal); PIVE coupling EM with an EIE lower bound as a second criterion, with a corrected-PIVE follow-up re-deriving the privacy proofs.	[17, 95, 96, 99, 106, 108]
(C′) Composition / dynamic / context-aware. *Cope with repeated releases or evolving environments.*	Regionalized DPIVE varying the EIE bound and budget by spatial cluster; Eclipse defense dynamically modulating protection against long-term observation attacks; dynamic-spectrum-sharing variants maintaining Geo-Ind during evolving auctions.	[97, 100, 101]
(D′) Output validity & density-aware adaptation. *Ensure outputs are admissible and informed by data distribution.*	Cooperative cloaking-region obfuscation for vehicular networks (CRO amortizes per-user budget across neighbors); Truncated EM (TEM) excluding low-density tail outputs; Perturbation-Hidden, which guarantees only *plausible* outputs (no vehicles in lakes) — empirically reduces the ∼50% off-manifold rate of vanilla Geo-Ind to 0%.	[94, 102, 142]
(E′) Theory, optimization integration, & system deployment. *Couple EM with LP / metric extensions / large systems.*	The canonical LP+EM hybrid (NHDM kernel + OPM solver); multi-objective evolutionary EM (MOEA) returning Pareto-optimal privacy-utility configurations; Lipschitz-extension generalizations to node-private graph statistics (subgraph counts, degree distributions); EM for graph-infrastructure obfuscation; condensed local DP for small-population datasets.	[3, 85, 98, 107, 109]

Table 9: OPM beyond the Bordenabe LP anchor (Sec. 3.3), organized by which ingredient of the LP recipe ($\min_Z \pi^\top (Z \odot c)\mathbf{1}$ subject to mDP constraints) is varied.

Recipe DOF	Variation & key technical idea	Representative works
(A″) Cost matrix $c_{x,y}$. *Replace geometric distance with task-specific cost.*	Task-completion-quality scores in MCS objectives; MINLP travel-distance for crowdsourcing matching, with multi-task-allocation extensions; multi-criteria objectives combining conditional entropy with worst-case quality loss; route-endpoint shifting cost that LP-optimizes which nearby POIs to use as proxies; pricing-aware ride-on-demand loss; group-based traffic-flow cost; multi-point "Geo-Ind masking" producing an obfuscation *set* rather than a single point.	[2, 113, 115, 116, 120, 122, 124, 135]
(B″) Constraint structure / scalability. *Tame the cubic constraint count of the canonical LP.*	Spanner-graph approximation; hierarchically well-separated trees that turn the LP into a cascade of small sub-LPs; Benders decomposition with dataset partitioning; Dantzig–Wolfe column generation specialized for road-network shortest-path metrics; peer-set constraint pruning for time-critical settings; graph-based CORGI approximations decoupling server-side optimization from user-side customization; shortest-path metric reformulation on road networks.	[32, 78, 110, 111, 114, 121, 123]
(C″) Cross-family hybrids. *Combine LP with closed-form HDM/NHDM kernels.*	Canonical LP+EM hybrid fixing far-away entries to a closed-form EM kernel and optimizing only near-neighbor entries (also covered as the canonical hybrid in Sec. 3.4); Bayesian remapping composing a closed-form HDM with an LP-derived post-processor that improves utility without re-paying privacy budget.	[3, 74]
(D″) Alternative optimization paradigms. *When LP is too rigid or the domain is too high-dimensional.*	Contract Theory under information asymmetry for cooperative localization; Stackelberg / bilevel games balancing accuracy against privacy; deep RL (A3C) for 3D-spatial task allocation; swarm/heuristic optimization for ride-on-demand.	[2, 89, 125, 126]
(E″) Theory & system integration. *Foundational analysis and large-system deployments.*	Dual-problem proof of HDM optimality for identity queries; metric-geometry characterization of the privacy-utility frontier; rigorous re-analysis of PIVE that identified privacy-proof flaws and proposed corrections; IoT proxy-gateway selection under DP for energy-efficient smart homes; traffic-flow-aware fake-trajectory selection.	[19, 96, 127–129]

Obfuscation methods	Features				
	HDM	NHDM	OPM	K	Utility loss definition
CCS 2012 [131]			✓	30	[G] Expected & max distance between real and reported locations
CCS 2013 [5]	✓			—	[G] Expected distance between real and reported locations
CCS 2014 [110]			✓	50	[G] Expected distance between real and reported locations
PETS 2015 (Privacy game) [132]			✓	300	[T] Expected utility cost (depends on applications)
PETS 2015 (Elastic metric) [92]		✓		—	[G] Expected distance between real and reported locations
ICDM 2016 [133]			✓	57	[G] Expected residual standard error
WWW 2017 [134]			✓	16	[G] Expected travel distance
NDSS 2017 [95]			✓	50	[G] Expected distance between real and reported locations
CCS 2017 [135]			✓	25	[T] Average & worst-case quality loss (depends on applications)
TIFS 2017 [87]	✓			—	[T] Task Utility (Satisfaction Ratio)
PETS 2017 [74][H]	✓		✓	—	[T] Expected quality loss (depends on applications)
CSF 2018 [136]	✓			30	[G] Expected distance between real and reported locations
TIFS 2018 [44]	✓			—	[T] Task Utility (Distortion/LBS Accuracy/Precision and Recall)
ICDM 2019 [26]	✓			100–300	[T] Author predictions/Accuracy
EDBT 2019 [78]	✓			100–300	[T] Expected utility cost
TMC 2020 [32]			✓	300–400	[G] Expected difference between real and estimated travel distance
PETS 2020 [83]	✓	✓		50	[G] Mean-squared error
WSDM 2020 [20]		✓		50–300	[T] Expected utility cost (depends on applications)
PETS 2020 [137]	✓			—	[G] Expected distance between real and reported locations
ACL 2021 [105]		✓		—	[T] Task Utility (Accuracy)
TDSC 2021 [116]	✓			—	[T] Expected travel distance/Acceptance ratio
CVPR 2021 [9]	✓			—	[T] Task Utility (Reidentification ratio)
ICDE 2021 [138]	✓			—	[T] Task Utility (Precision/Recall)
TWC 2022 [80]	✓			—	[G] Average expected error between the perturbed location and the actual location
SIGSPATIAL 2022 [128]				100	[G] Expected difference between real and estimated travel distance
UAI 2022 [3][H]		✓	✓	400	Expected utility cost (depends on applications)
IoTJ 2022 [89]	✓		✓	—	[T] Task Utility (type-k cooperative nodes' utility)
IoTJ 2022 [59]			✓	—	[T] Task Utility (MSE)
CCS 2022 [108]		✓		—	[T] Task Utility (Accuracy of range queries)
TMC 2023 [99]		✓		—	[G] Expected distance between real and reported locations
EDBT 2023 [114]			✓	70	[G] Expected distance between real and reported locations
TKDE 2023 [37]		✓		300	[G] The difference between the obfuscated and original covariance matrices
EDBT 2024 [123]			✓	300–400	[G] Expected distance between real and reported locations
IJCAI 2024 [111]			✓	1000	[T] Expected data quality loss
PETS 2025 [140][H]		✓	✓	1,600	[G] Expected difference between real and estimated travel distance
CCS 2025 [141]			✓	5,000	[G] Expected difference between real and estimated travel distance

Table 10: Representative mDP mechanisms surveyed in this paper. Columns indicate, respectively: (i) the venue and citation of the work; (ii)–(iv) whether the work uses a *homogeneous distance mechanism* (HDM), a *non-homogeneous distance mechanism* (NHDM), and/or an *optimized perturbation mechanism* (OPM) — multiple checkmarks (and the [H] marker on the paper name) denote *hybrid* designs that span two categories (cf. Sec. 3.4); (v) the size K of the discretized input/output domain used in the experiments reported in the cited paper (this is the cardinality that controls the LP size for OPM and the candidate-set size for NHDM; "—" indicates that the work does not fix a finite domain size — e.g., the planar Laplace kernel is defined on the continuous plane); and (vi) the *primary* utility-loss definition reported by the authors, tagged [G] for *geometric distortion* (e.g., expected ℓ_2 error, mean-squared error, expected travel-distance error) or [T] for *task-driven utility* (e.g., classification accuracy, retrieval precision, satisfaction ratio), per the discussion in Sec. 3.4.

Dataset	Type	Application Domain	Size	Representative Works
GeoLife [147]	Location	Trajectory obfuscation, semantic-aware perturbation	17,621 trajectories / 182 users	[23, 44, 45, 148]
T-Drive [149]	Location	Vehicle routing, trajectory sanitization	10,357 taxis (∼15M GPS points)	[44, 62, 65]
Gowalla / Brightkite [150]	LBSN	Location privacy, spatial crowdsourcing	G: 196k users; B: 58k users	[33, 138, 151]
Foursquare [152]	Check-in	POI recommendation, friend matching (LDP)	800k check-ins (NYC/Tokyo)	[35, 153]
Cabspotting/ Portocabs [154, 155]	Location	Urban mobility modeling, map-matching attacks	441–500 taxis (1 month–1 year)	[148]
MovieLens [156]	Ratings	Collaborative filtering, embedding perturbation	Up to 25M ratings, 162k users	[39]
Yelp [157]	Text/Geo	LBS privacy–incentive tradeoff	Millions of reviews / users	[64]
EMNIST [158]	Image	FL with group privacy	814k characters / 3.5k writers	[90]
IMDb [159]	Text	Text privatization, sentiment masking	50k labeled + 50k unlabeled	[84]
PAN11 / PAN12 [160, 161]	Text/Email	Author obfuscation, stylometry	Varying corpus per year	[26]
NLP Benchmarks [162, 163]	Text	Embedding-level privatization	MR: 10.7k; CR: 3.8k samples	[26]
US Cities / Census [164, 165]	Tabular/Geo	Histogram queries, policy-aware DP	>40k cities, millions of samples	[31, 83]
Twitter Geo [166]	Social/Count	Spatial DP for social activity	Stream data (API-dependent)	[31]
Graph Datasets [167]	Graph	Node/edge perturbation, link privacy	e.g., Cora (2.7k nodes)	[31]
OpenStreetMap [169]	Map	Road network perturbation, topology analysis	Global, dynamic POI data	[74]
GRUMPv1 [170]	Census/Grid	Population density modeling	Global grid (1km resolution)	[77]
Synthetic Datasets	Simulated	Controlled mDP benchmarks (grid/LBS)	Task-dependent	[76, 123, 172]

Table 11: Representative datasets commonly used in mDP research.