Differential Privacy Configurations in the Real World: A Comparative Analysis

Michael Khavkin, *Member, IEEE,* and Eran Toch, *Member, IEEE*

***Abstract*—An increasing number of technologies depend on the large-scale collection of individual-level data, whether for gathering statistical insights from billions of users or training AI models. However, reliance on personal data raises privacy concerns that, in turn, limit the collection and analysis essential to these technologies. Differential Privacy (DP) has gained traction in both academia and industry, ensuring privacy by adding carefully crafted noise to data or its outputs based on a pre-defined DP parameter ε. As real-world implementations emerge, we can examine how DP is practically used beyond academic settings, supporting industry adoption and expanding knowledge on DP applications. Using a systematic process, we comprehensively surveyed the deployed parameters of DP configurations in both commercial and governmental implementations ($n = 140$) and compared them to those employed in academic research. We also propose a high-level taxonomy for DP configuration, capturing practical implementations of differentially private Machine Learning (ML) and Federated Learning (FL) applications, highlighting key factors, including the privacy unit and ε. Our results show that, on average, ε values utilized in the industry span a wider range than those in academic research, with distinct configuration policies for governmental and commercial organizations. Moreover, we identified contrasting reasoning behind ε selection across deployment environments, alongside insufficient transparency in industry disclosures of DP parameters and limited support for user-oriented configuration. Finally, we discuss how the collected knowledge can be used to create methodological guidelines for the configuration of DP in real-world environments, supporting the vision of an Epsilon Registry.**

***Index Terms*—Differential Privacy, Privacy Budget, Privacy-Preserving Machine Learning, Federated Learning, Survey**

I. Introduction

DATA has become an essential resource for AI-driven information systems, significantly advancing data analysis, particularly Machine Learning (ML) models, across various domains such as healthcare, social networks, and smart energy. The data used to develop such models and use them for analysis often includes sensitive individual-level information, which may pose a threat to privacy. Differential Privacy (DP) [1] has been proposed as a rigorous privacy guarantee for computation mechanisms, under which an external party cannot infer with a high probability individual-level information. The level of protection offered by differential privacy is primarily determined by the privacy loss parameter ε and the parameter δ, which bounds the probability $1-\delta$ with which the privacy guarantee holds. These parameters together quantify the privacy loss incurred when personal data is used in a differentially private analysis, effectively capping the amount of analysis performed on that data to minimize the risk of revealing sensitive information. To satisfy DP in a data analysis, DP mechanisms inject carefully calibrated noise (proportional to the privacy budget ε [1]) into the data or its derivatives, thereby masking individual contributions. In machine learning, noise is typically added to model parameters, such as weights, or to gradient computations in the case of deep learning neural networks. For instance, a medical institute reporting the number of patients with a rare disease adds a small amount of random noise to the count, preventing adversaries from determining whether a specific person is included in their dataset [2]. Similarly, a company publishing the average employee salary introduces noise to the reported value, ensuring that individual salaries remain undisclosed even when different results on different data versions are compared [3].

Due to its strong, future-proof, and adversary-agnostic guarantees [4], [5], differential privacy has gained growing interest among industry organizations. Many have begun incorporating DP mechanisms into products and machine learning models to provide formal privacy protection [6]–[8]. Notable examples include next-word prediction in Google's English Gboard app [9] providing $(\varepsilon = 4.79, \delta = 10^{-10})$-DP protection and Apple's Photos app, which uses $(\varepsilon = 1, \delta = 1.5 \times 10^{-7})$-DP when selecting key photos for iOS apps, such as Memories and Places. Recently, several applications emerged as a result of the COVID-19 pandemic, including the release of $(\varepsilon = 2.19, \delta = 10^{-5})$-differentially private trends in Google search data on vaccinations [10], and $(\varepsilon = 2.64)$-DP mobility patterns across regions to help researchers understand the pandemic's societal impact [11]. Special attention has also been given to the use of DP in governmental services [12]. A notable milestone was the U.S. Census Bureau's adoption of DP in its Disclosure Avoidance System (DAS) [13], which demonstrated the practical viability of DP and triggered further research by other National Statistical Offices seeking to apply it in their own census releases [14], [15]. DP has also been deployed in various other domains, including media [16], communications [7], [17], and smart energy [18].

Although introduced almost two decades ago, the transition of DP from theory to practice has raised several practical questions. These questions reflect the complexity of implementing differentially private mechanisms for data analysis, often stemming from a limited understanding of how to operationalize DP and the lack of concrete implementation guidelines, which can all lead to misconfiguration. The term *DP configuration* is considered herein as the process during which data practitioners define the DP protection guarantees

Manuscript received March 4, 2025; revised July 4, 2025. This work was supported in part by a grant from the Tel Aviv University Center for Artificial Intelligence and Data Science (TAD).

Michael Khavkin and Eran Toch are with the School of Industrial & Intelligent Systems Engineering, Tel Aviv University, Tel Aviv, Israel (email: khavkin1@mail.tau.ac.il; erant@tauex.tau.ac.il).

This article has been accepted for publication in IEEE Transactions on Knowledge and Data Engineering. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TKDE.2025.3603731

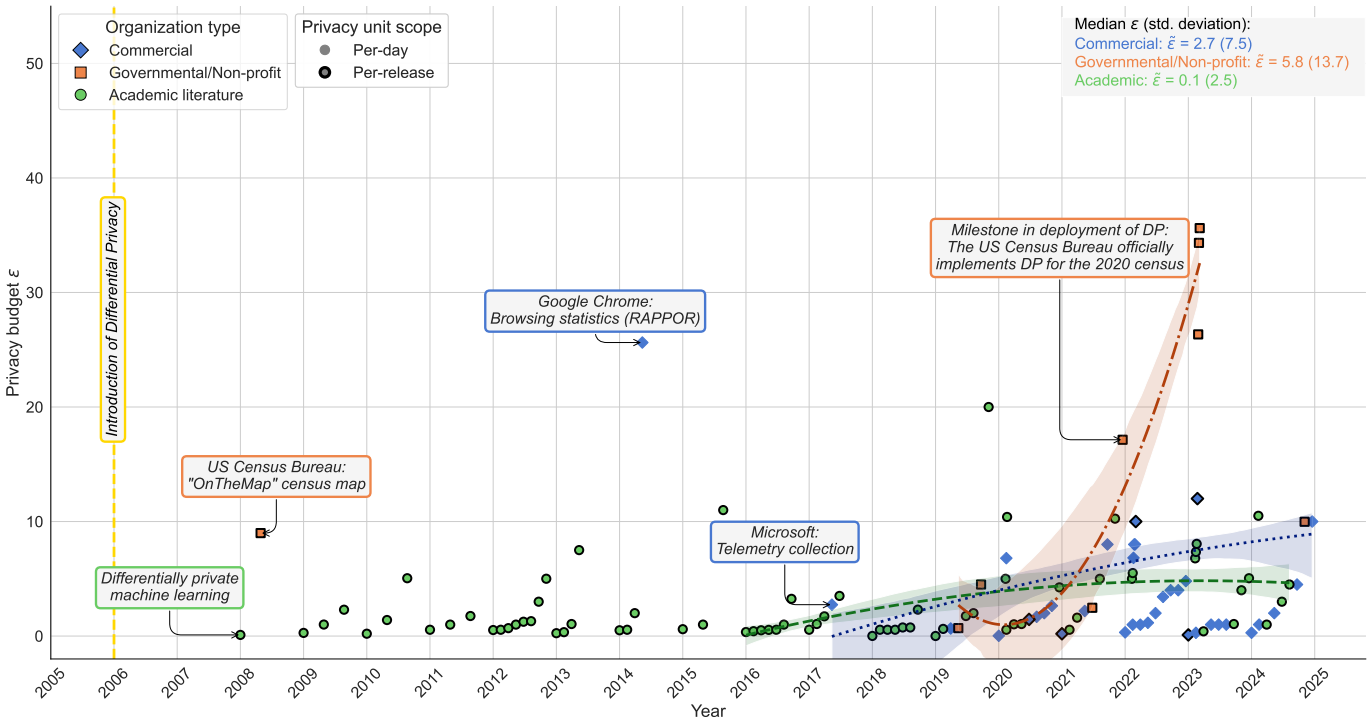

Fig. 1. Evolution of the privacy loss budget ε used in academic, commercial, and governmental/non-profit DP deployments. Academic research configurations are not exhaustive and serve as a baseline for comparison. Privacy parameters of use cases analyzing data temporally were converted to a common ε per day units (when applicable). For static data releases (e.g., Census), the original one-time ε is plotted without normalization. δ is not shown due to its low and indistinguishable values across all deployments. In academic literature (which typically evaluate a range of values) median ε is shown. Dashed lines depict a forecast through a polynomial regression line fitted on the ε values for each organization type based on the last 5 years.

their system will offer data subjects, design the setup under which these guarantees are met and allocate the required privacy budget ε to that end. In practice, in the case of an already deployed DP system, the configuration may only refer to the setting of ε (and δ, if applicable). Choosing the necessary privacy budget for a specific data analysis is viewed as a decision-making process rather than a strictly technical statistical decision, primarily driven by a social choice and privacy policies that consider potential privacy risks [19]. Furthermore, similarly to other Privacy-Enhancing Technologies (PETs), the use of DP for privacy protection comes at the cost of a decrease in data utility—a property that has become the focal trade-off in research concerning privacy, and particularly DP, referred to as the privacy-utility trade-off [20]. Hence, the tension between privacy and utility has spurred a debate about the responsibility of data practitioners in considering this trade-off because misconfiguration of ε can even become life-threatening, as in the case of DP algorithms for guiding pharmacogenomic dosing [21]. Misconfiguration of DP can also lead to a potential overestimation of privacy protection caused by data practitioners' misconception that their systems offer the desired DP protection, when, in fact, users can still be susceptible to privacy harm.

A line of works [22], [23] survey DP with a profound formal background for its privacy guarantees. Other survey articles examine the use of DP in building privacy-preserving analysis mechanisms across various domains, including healthcare [24], social networks [25], communication systems [26], and geo-analytics [27]. A special emphasis was placed on surveying differentially private techniques for training and deploying machine learning models [28], [29], which are widely used in everyday products, such as smartphone keyboards and text messaging apps. These also include the use of Federated Analytics (FA) [30] and Federated Learning (FL) [31], which extend traditional centralized ML to distributed settings where multiple parties can collaboratively train ML models while ensuring DP. However, the existing surveys focus on a description of empirical research on DP applications and do not cover the exact implementation details of DP in deployed commercial or governmental products. Analyzing the way existing commercial, governmental and academic DP implementations are configured is crucial to understanding how organizations can set up DP as part of their operational process and the potential limitations and promises of their deployments. In addition, to the best of our knowledge, existing surveys do not offer a comprehensive account of deployed privacy configurations in real-world DP deployments, including the unit of privacy protection, its granularity, and the underlying DP mechanism.

In this paper, we conduct a systematic literature review on practical DP configurations utilized for commercial and governmental use cases in the industry. We give special attention to the gap between the configuration of DP in the industry and that established for academic research, which can naturally diverge. Understanding the gap between commercial or governmental deployments and academic research may help

in understanding how the transition from theory to practice can be made more accessible to data practitioners working on the deployment of DP mechanisms, thereby expanding the existing body of knowledge on DP. To that end, we present a comprehensive list of recent DP deployments across various use cases, detailing their configurations and illustrating the distribution of the ε parameter (Figure 1).

Through our systematic literature review, we highlight several insights about the challenges in the configuration of DP and suggest potential research directions for mitigating them. Our analysis reveals that governmental organizations, such as the U.S. Census Bureau, tend to allocate privacy budgets across a broader range, while commercial companies generally operate within a narrower range of ε values. In contrast, academic research has typically used much lower ε values, often $\varepsilon < 1$, to prioritize privacy and demonstrate the practical viability of differential privacy methods. Moreover, our survey identified that the selection of DP parameters in both industry and academia is driven by a range of factors, including arbitrary choices, community best practices, utility-based tuning, and regulatory requirements. Our survey can help data practitioners of different roles, including data scientists and privacy stewards who are responsible for the enforcement of privacy policies, understand the rationale behind existing DP configurations. Additionally, our findings can serve as a reference for data practitioners to establish a baseline for their configurations and position their use cases within the landscape of industry-deployed solutions. This can help in promoting the operationalization of DP, thereby breaking through the barrier associated with practical DP implementation and promoting Dwork's et al. vision of an "Epsilon Registry" [19].

To summarize our contributions, we aim to answer two key questions as follows.

1) How do DP configurations, particularly the (ε, δ) parameters, vary across commercial deployments, governmental applications, and academic research?
2) How do the underlying rationales for selecting DP configurations differ among commercial deployments, governmental applications, and academic research?

This paper is organized as follows. Section II reviews the formal guarantees of differential privacy (DP). Section III details our systematic literature review methodology. Section IV presents a taxonomy of key DP configuration factors. Sections V and VI examine and compare DP applications and configurations in industry and academia. Section VII concludes with a discussion of key findings.

II. Background and Preliminaries

A. Differential Privacy (DP)

Perturbation-based privacy methods have been proposed to address the vulnerability of syntactic privacy protection methods. The key idea behind such methods is to introduce a randomized perturbation that protects the privacy of individuals whose personal information is shared with a data processor. Built on this notion, **Differential Privacy (DP)** [1] has been proposed as a mathematically rigorous privacy guarantee. Let $\mathcal{D} = (x_1, \ldots, x_n)$ be a dataset where each x_i represents the

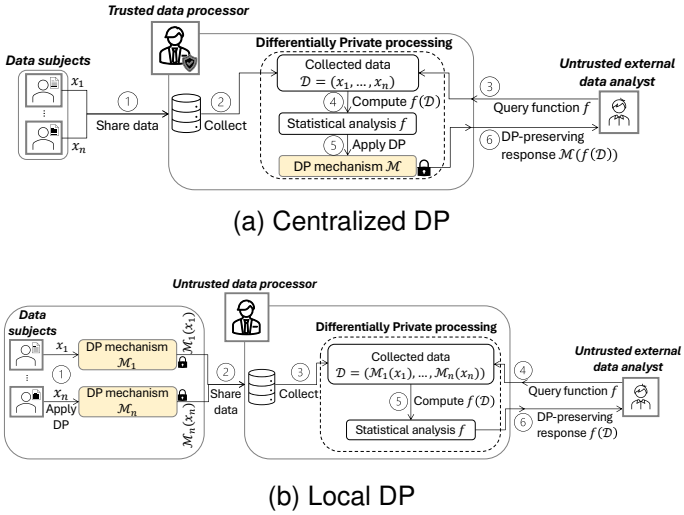

Fig. 2. Typical flows of an interactive data analysis under centralized and local DP settings. The main actors are depicted in gray boxes.

data contributed by data subject i. A randomized mechanism $\mathcal{M} : \mathcal{D} \to \mathcal{S}$ that performs a statistical analysis f satisfies (ε, δ)-differential privacy, for $\varepsilon > 0$ and $\delta \in [0, 1]$, if and only if for all pairs of neighboring datasets $D, D' \in \mathcal{D}$ differing in at most one record, and for all output subsets $S \subseteq \mathcal{S}$,

$$\Pr[\mathcal{M}(D) \in \mathcal{S}] \leq \exp(\varepsilon) \cdot \Pr[\mathcal{M}(D')] \in \mathcal{S}] + \delta \qquad (1)$$

The additive term δ accounts for the probability that ε-DP is violated, implying that ε-DP is satisfied with probability $1-\delta$. When $\delta = 0$ the mechanism satisfies *pure* ε-DP, whereas $\delta > 0$ corresponds to *approximate* (ε, δ)-DP.

The parameter ε controls the maximum amount of information that can be learned about any individual's data through the analysis of a dataset. In other words, the outcome of an analysis is equally likely, up to a multiplicative factor e^{ε}, independent of whether any individual's data is present in the analyzed dataset. The key idea behind the privacy loss budget parameter is that privacy can be viewed as a resource, which is consumed during the course of a data analysis, until all "privacy" is exhausted and no additional analysis can be performed on the data. Hence, the lower the privacy loss budget ε, the higher the protection DP offers, and vice versa. The failure probability δ is conventionally set to a negligible value, often based on the heuristic $\delta < 1/n^2$, proportional to the dataset size n [5]. The allocation of ε and δ lacks clear guidelines, as it is often regarded as a policy-driven process, rather than a pure technical procedure.

B. Local Differential Privacy (LDP)

DP can be implemented in either a centralized mode (i.e., *centralized DP*) or a local mode (i.e., *local DP*) [5], [32], as illustrated in Figure 2. In centralized DP, data is first aggregated at the data processor's side before DP protection is applied on the output of an underlying data analysis. The centralized approach assumes that data subjects trust the data processor, allowing it to store raw sensitive data. In contrast, Local Differential Privacy (LDP) [32] has been proposed as

a DP mechanism that is suitable for use cases where the data processor is deemed untrusted, aligning with current real-world scenarios. In local DP, DP protection is first applied locally at the data owner's side before sending the masked data to the data processor. Consequently, the data processor receives a privacy-preserving version of the data, rather than the raw data, ensuring stricter privacy compared to the centralized mode. Formally, a randomized mechanism $\mathcal{M} : \mathcal{V} \to \mathcal{S}$ is ε-local differentially private if and only if for any pair of inputs $v, v' \in \mathcal{V}$, and for any possible output $y \in \mathcal{S}$ of $\mathcal{M}$,

$$\Pr[\mathcal{M}(v) = y] \leq \exp(\varepsilon) \cdot \Pr[\mathcal{M}(v')] = y] \tag{2}$$

LDP has also laid the foundation to personalized approaches to satisfying DP, such as Personalized Differential Privacy (PDP) [33], where the privacy guarantees are set locally at an individual level, setting a distinct DP budget ε for each individual according to their privacy preferences. One disadvantage of LDP is that it requires adding more noise than in centralized DP to achieve the same level of protection, with negative consequences on data utility [5]. Moreover, when users contribute numerous correlated data points (e.g., movie views), adding noise to each event may be insufficient to protect overall privacy, thereby reducing the effectiveness of local DP at the user level [34]. Accordingly, *hybrid DP* approaches, such as the shuffle model [35], have been introduced to combine the benefits of local and central DP.

C. Differentially Private Perturbation Mechanisms

Differential Privacy is typically implemented through noise-addition mechanisms that introduce carefully calibrated noise to data, analysis outputs, or intermediate computations, to limit the leakage of individual information. The Laplace mechanism [1] or the Gaussian mechanism [5] are commonly used to satisfy DP for numerical outputs by drawing noise from Laplace and Gaussian distributions, respectively. The geometric mechanism [36] was later proposed as a discrete variant of the Laplace mechanism. Recently, the discrete Gaussian mechanism [37] was introduced as a practical alternative to the continuous Gaussian mechanism, offering comparable privacy and accuracy guarantees while improving interpretability in discrete data settings, such as the release of census counts.

When the protected output is categorical, or when a data processor needs to select a privacy-preserving answer from a finite or infinite set of possible outputs, the exponential mechanism [5] can be used. The exponential mechanism can be replaced with the Report Noisy Max mechanism [5] in case of a finite set of answers. The main advantage of the exponential mechanism is that the privacy cost of the mechanism is ε regardless of the size of that set because it releases only the candidate answers with the largest noisy value. In the case of a stream of queries, the Sparse Vector Technique (SVT) [5], including its basic form known as the Above-Threshold algorithm, can be employed. A summary of the basic perturbation mechanisms is provided in Table I.

TABLE I
CHARACTERISTICS OF BASIC PERTURBATION MECHANISMS. CENTR.=CENTRALIZED; NUM.=NUMERIC; CAT.=CATEGORICAL; BOOL.=BOOLEAN.

Mechanism	DP Mode		Data Type			Guarantee
	Local	Centr.	Num.	Cat.	Bool.	
Randomized Response [38]	●	○	○	●	●	$\ln(3)$-DP
Laplace [1]	●	●	●	○	○	ε-DP
Gaussian [5]	●	●	●	○	○	(ε, δ)-DP
Geometric [36]	●	●	●	○	○	ε-DP
Exponential [3]	○	●	○	●	●	ε-DP
Report Noisy Max [5]	○	●	●	●	●	ε-DP
Sample & Aggregate [39]	○	●	●	●	●	(ε, δ)-DP
Functional [40]	○	●	●	○	○	ε-DP
Above Threshold [5]	○	●	●	○	●	ε-DP
Sparse Vector Tech. [3]	○	●	●	○	●	(ε, δ)-DP
Propose-Test-Release [41]	○	●	●	○	○	(ε, δ)-DP

D. Relaxations of DP

Several practical relaxations of DP [42]–[45] model privacy loss as a random variable and bound its distribution, capturing average rather than worst-case guarantees through divergence-based definitions. The most common definition for divergence is Rényi divergence [46], which measures the closeness $D_\alpha(P||Q)$ of two probability distributions P and Q over $\mathcal{R}$. Rényi DP (RDP) [42] was proposed based on that metric as a relaxation of DP, which avoids the definition of "catastrophic" failure while preserving all the composition properties of DP. Formally, a randomized mechanism $\mathcal{M} : \mathcal{D} \to \mathcal{S}$ is (α, ε)-RDP for some order α if for any pair of neighboring datasets $D, D' \in \mathcal{D}$, their Rényi divergence satisfies

$$D_\alpha(\mathcal{M}(D) \| \mathcal{M}(D')) \leq \varepsilon. \tag{3}$$

A key advantage of Rényi DP is its versatility: any mechanism satisfying (α, ε)-RDP also satisfies $(\varepsilon + \log(1/\delta)/(\alpha - 1), \delta)$-DP for any $0 < \delta < 1$, simplifying configuration and enabling easier comparison across algorithms.

DP was later formulated as Zero-Concentrated DP (ρ-zCDP) [44], defined in terms of Rényi divergence with a stronger requirement than RDP, limiting the privacy parameter to a single parameter ρ, which controls the expectation and standard deviation of the privacy loss. Similarly to RDP, the obtained guarantees of ρ-zCDP can be converted back to (ε, δ)-DP, such that a ρ-zCDP mechanism provides $(\rho + 2\sqrt{\rho \log(1/\delta)}, \delta)$-DP for any $\delta > 0$ (Proposition 1.3 in [44]). Both RDP and the concentrated-DP versions generate noise based on the Gaussian mechanism [5]. Truncated CDP (tCDP) [45], a relaxation of CDP, loosens the requirement of Gaussian-concentrated privacy loss, requiring the noise distribution to be only sub-exponential in its tails. tCPD also allows for privacy amplification via subsampling [47], yielding exponentially more accurate analyses.

The advantage of RDP, zCDP, and tCDP lies in their ability to yield lower ε values (indicating stronger privacy guarantees) for the same level of added noise, thereby enabling more analyses within a given ε budget. The relaxed constraints of zCDP, along with its improved privacy accounting, have contributed to its widespread adoption in real-world applications, particularly differentially private distributed machine learning, including Federated Learning [48].

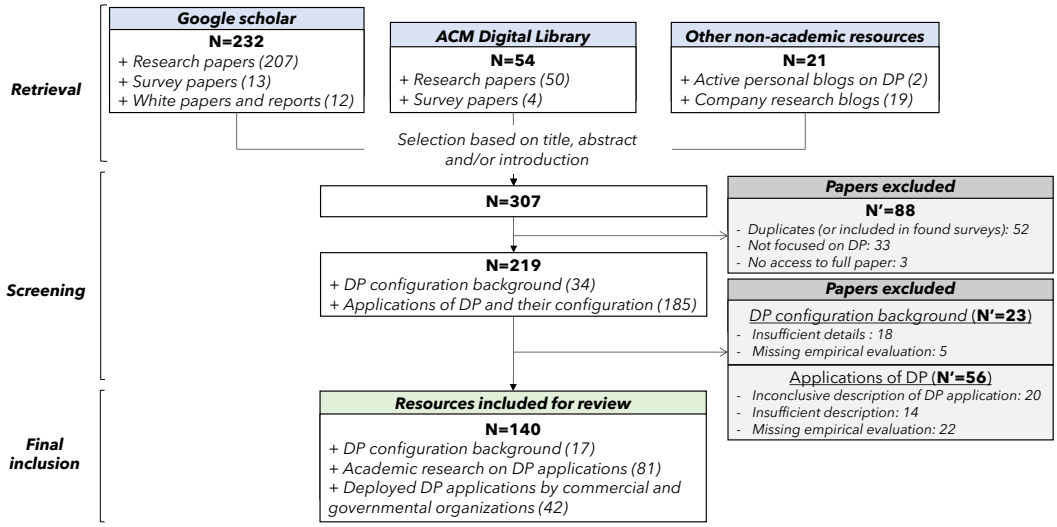

Fig. 3. Systematic literature review flow.

E. Differentially Private Machine Learning

Although DP was initially associated with statistical analysis, such as histograms and descriptive statistics (e.g., count, average, and median), modern real-world applications predominantly use it in Machine Learning (ML) for model training. In the context of ML, DP operates by injecting noise into the derivatives of the model's training through a randomized mechanism. For instance, in regression models [40], [49], DP is enforced by adding noise to the objective function's coefficients. In classification models, such as Gaussian Naive Bayes [50], a differentially private mechanism perturbs the learned means and variances used to compute conditional probabilities.

Recent advances in large-scale machine learning have increased the demand for scalable training methods that preserve both privacy and utility. These models often rely on Deep Neural Networks (DNNs), trained using Stochastic Gradient Descent (SGD), which iteratively updates the network's weights to minimize the model's objective function. To preserve differential privacy, noise is added to gradients during training via the Differentially Private Stochastic Gradient Descent (DP-SGD) algorithm [51], a method widely adopted in commercial applications. DP-SGD preserves DP in each training iteration by first clipping each instance's gradients to a fixed norm, then adding Gaussian noise to the average per-instance gradients. A central aspect of DP-based ML training is privacy accounting, which tracks cumulative privacy loss across iterations to ensure it stays within a predefined loss budget [51].

Federated Analytics (FA) [30] enables large-scale, privacy-preserving data analysis by allowing distributed analytical queries, such as computing averages, across multiple parties without exposing raw data to a central entity. Federated Learning (FL) [31] extends this concept to machine learning by treating model training as a distributed optimization problem. Unlike centralized training, FL keeps data decentralized, with clients training models locally on their own data and sending only model updates to a central server, which iteratively aggregates them into a global model. This paradigm is well-suited for resource-constrained IoT devices, such as smartphones. Differentially Private Federated Averaging (DP-FedAvg) [52] is a common approach for integrating DP into FL, providing user-level DP guarantees [51]. It extends the DP-SGD method [51] by performing a few local training steps based on each user's private data before the global model update is computed by clipping and averaging the local updates, followed by the addition of noise to the global update.

III. Systematic Literature Review Methodology

To create an initial set of articles for inclusion, we utilized Google Scholar to locate research articles from all types of venues, including conferences, journals, technical reports, and workshops. In addition, the ACM Digital Library was used to locate relevant articles, especially from computing journals and conferences. All articles were retrieved based on their titles, abstracts, and their introduction sections. We did not restrict our search to specific publishing time frames to maximize the number of search results, though all articles were naturally published after DP's introduction in 2006. However, since the adoption of DP has only recently begun to gain traction in the industry, the majority of reviewed articles describing real-world use cases of DP have been published after 2018. We searched for queries containing the exact phrases '*Differential Privacy*' (or '*Differentially Private*'), combined with keywords related to configuration (e.g., '*epsilon*,' '*privacy budget*,' '*parameters*') or practical deployment (e.g., '*practical*,' '*practice*,' '*deploy*,' '*real-world*,' '*applications*,' '*industry*'). Figure 3 depicts the flow of our systematic review process.

We expanded our search to find other relevant articles by examining the reference lists in the retrieved articles. We excluded articles that did not describe general methods for the selection of ε or had insufficient empirical evaluation. Despite the variety of recent applications and use cases adopting DP in the real world, not all could be utilized for our review due to

incomplete documentation or a lack of methodological details on how DP was implemented.

In cases where articles that pertained to commercial deployments did not include any implementation details but provided adequate motivation for the underlying use case, we included these in our survey to obtain an understanding of the transparency of organizations implementing DP. In other cases, recently deployed products incorporating DP could only be found online through a Google search and not in academic search engines because of unpublished documentation to support them. To address this issue, we also searched for online blogs (2 blogs [53], [54]) that directed us to the sources where recent use cases of DP are described, in addition to official research blog posts of companies in the industry (e.g., Google's or Microsoft's research blogs). Our sample consisted of 140 DP-related articles, serving as input to our survey across four topics: general background of DP configuration (17 articles), academic research on DP applications (81 articles, used for comparison), and DP deployments by commercial and governmental organizations (42 articles).

IV. High-level Taxonomy for Differential Privacy Configuration

Configuring DP for data analysis is a multi-faceted process that typically involves defining the privacy parameters ε and δ. While these parameters lie at the core of the configuration, commercial and governmental deployments often require additional preparatory steps to accommodate practical constraints and align the setup with specific use case objectives. However, with the growing availability of black-box DP tools, many of these complexities are mitigated. In such cases, data administrators often rely on default built-in algorithms, primarily configuring only the ε and δ parameters.

To better understand the configuration process and contribute to the existing body of knowledge on DP operationalization [5], [19], [28], we categorized key conceptual factors involved in DP configuration into two main groups: privacy guarantees and operational design (Figure 4), as follows.

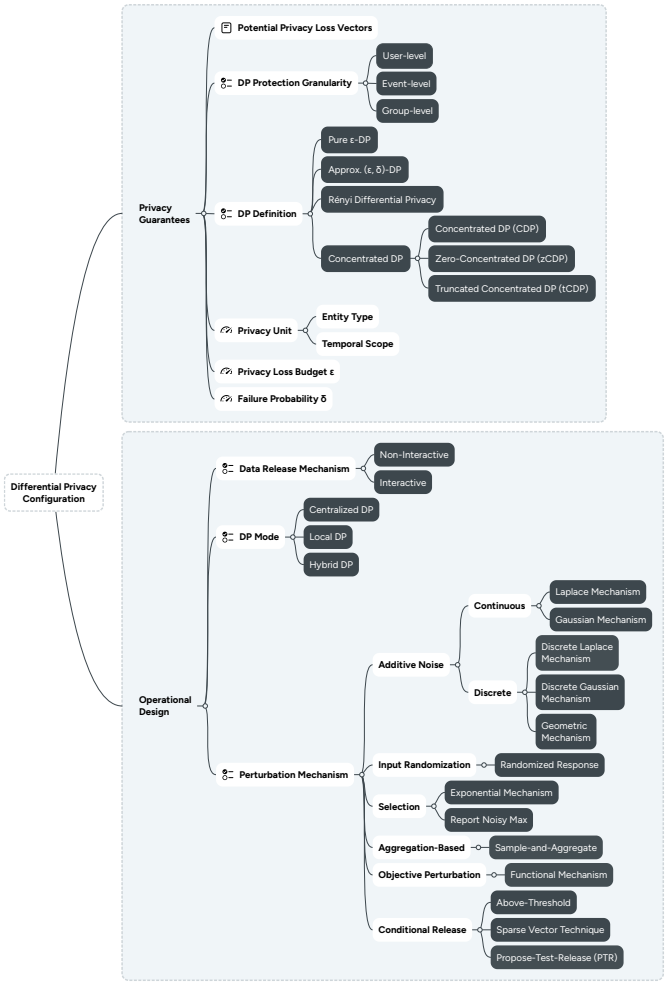

Fig. 4. Taxonomy of key factors in the process of DP configuration, comprised of tunable settings and design choices. Icons depict abstract ( ), tunable ( ), and categorical ( ) factors. Only the fundamental building blocks of perturbation mechanisms are shown, forming the basis of DP methods.

A. Differential Privacy Guarantees

First, the data processor (alternatively, the organizational data steward/controller) evaluates potential avenues through which sensitive information may be leaked, leading to privacy loss [19]. Privacy loss is incurred when collected data is used in ways that affect others' experiences or reveal information about them. The analysis of potential paths for privacy loss can help data practitioners who are responsible for operating and configuring DP systems to get a high-level understanding of the constraints that should be considered in the process and hence can impact the setting of ε. Then, the data processor specifies the granularity of the guaranteed DP protection and the corresponding privacy unit that is protected by it. DP protection granularity refers to the level at which a differentially private mechanism protects from inferring private information. *User-level* protection granularity offers an ε-DP guarantee over all data instances associated with a user or device (e.g., a user's watch history), while *group-level* protection generalizes this to arbitrary groupings. In contrast, *event-level* protection (also referred to as instance-level protection) provides ε-DP for each individual data instance (e.g., a single movie watched), resulting in weaker overall protection for users with multiple events. The privacy unit [5], [19] defines the granularity of protection by bounding the cumulative privacy loss for a protected entity over a given time period. For example, in a streaming service, where users contribute viewing data daily, the privacy unit may be defined as one user per day to bound their privacy loss. Finally, the allocation of the privacy loss budget and the choice of failure probability are key factors determining the strength of the privacy guarantee and the number of allowable data analyses on a dataset.

B. Operational Design

Once the privacy unit is defined, the data processor can select the data release mechanism and the mode under which the mechanism operates. This may take the form of a non-interactive release, publishing differentially private data derivatives, or an interactive querying system that provides differentially private outputs to external analysts during inference (e.g.,

TABLE II
TAXONOMY-BASED MAPPING OF CONFIGURATION COMPONENTS IN GOOGLE'S DIFFERENTIALLY PRIVATE NEXT-WORD PREDICTION SYSTEM FOR A SINGLE-DEVICE USER TYPING IN U.S. ENGLISH.

Type	Component	Level	Explanation
Privacy Guarantees	Potential privacy loss vectors	• *Memorization of unique user phrases during training*: On-device training can still lead to memorization of unique phrases, which may be reconstructed from the final model. • *Leakage during local update aggregation*: Individual device contributions may be exposed or inferred by the central server during aggregation. • *Client re-identification during update propagation*: Frequent participation of a user's device in training rounds with insufficient time gaps may enable client re-identification.	
	DP protection granularity	User-level	Privacy guarantees apply to all data from a single user device.
	DP definition	zCDP	Uses Zero-Concentrated DP, which can be converted to (ε, δ)-DP.
	Privacy unit	User device per 24 hours	The entire update from a user's device is protected by DP. Participation is constrained by a system-enforced timer (e.g., 24-72 hours).
	Privacy budget ε	4.799	$\rho = 0.250$-zCDP is used given $\delta = 10^{-10}$, which corresponds to $(\varepsilon = 4.799, \delta = 10^{-10})$-DP.
	DP failure proba. δ	10^{-10}	
Operational Design	Data release mechanism	Interactive	Differentially private predictions are generated in response to user inputs.
	DP mode	Central DP	Noise is added on the trusted server side after aggregating local updates.
	Perturbation mechanism	Discrete Gaussian Mechanism	Differentially private optimization is performed via the DP-FTRL algorithm [55], which adds noise using the *Discrete Gaussian Mechanism*.

statistical estimates or model predictions). The release policy can then be used to determine whether DP will be provided via a centralized, local, or hybrid setting (e.g., federated analytics or federated learning).

Finally, the DP perturbation mechanism is determined based on the analysis type (either general statistical analysis or machine learning) and the data type (e.g., numerical, categorical or mixed) [19]. We also differentiate between centralized ML, where a single entity trains the model using all the raw data, and distributed ML (including Federated Learning), where multiple parties collaborate in the training process. Despite their structural differences, both approaches achieve DP through similar techniques, primarily by injecting noise either at the model's prediction level (using the sample-and-aggregate framework [39]) or during training. In the latter case, noise can be introduced at different stages, such as injecting noise to the model's weights, modifying the objective function using the Functional Mechanism [40], or perturbing the gradients in gradient-based models through algorithms, such as DP-SGD [51] and DP-FedAvg [31].

C. Example of Taxonomy-Based Configuration

To illustrate how our DP configuration taxonomy can help practitioners benchmark their deployments against other commercial implementations and identify trade-offs in privacy guarantees, we analyze Google's Next-Word Prediction (NWP) feature in the Gboard app (Google's virtual keyboard) [56]. Google employs machine learning to enhance the typing experience by predicting the next word a user is likely to type. This feature provides real-time word suggestions after each entry, enabling users to type more quickly and efficiently. Google employs Federated Learning to train its language model directly on users' devices without centrally collecting raw personal information, aggregating privately computed updates via a secure protocol that prevents the server from accessing any individual's data.

Table II maps the deployment's configuration decisions using our taxonomy (Figure IV), providing a structured view of the resulting privacy guarantees. This structured view also allows data practitioners to systematically compare the configurations of different deployments. For example, Gboard's configuration offers users with moderate DP guarantees, i.e., stronger than those provided by Recurve for smart energy metering [18] ($(6.8, 4.08 \times 10^{-8})$-DP), but weaker than those in Google Shopping's page-view count release ($(1, 10^{-9})$-DP) [57]. These differences also originate from distinct DP definitions: while both Gboard and Recurve provide central DP with a per-user daily budget, Google applies zCDP with a discrete perturbation mechanism, whereas Recurve adopts approximate DP with noise drawn from a continuous distribution.

V. REVIEW OF COMMERCIAL AND GOVERNMENTAL DP DEPLOYMENTS

In this section, we review both commercial and governmental DP deployments by multiple organizations across various domains, focusing on their use cases and DP configurations. In Table IV, we provide the list of properties for each use case with its exact reported DP parameters [1]. We present the configurations on a unified grid in Figure 1 to illustrate the distribution of varying ε values (a full list of depicted papers is provided in Table V in the Appendix).

A. Web, Communication & Browsing Behavioral Analytics

DP has been predominantly applied to large-scale behavioral analytics systems to enable privacy-preserving collection of user interaction data, such as web activity and browsing behavior. Google's initial deployment of DP was performed through the RAPPOR (Randomized Aggregatable Privacy-Preserving Ordinal Response) mechanism, used to collect private browsing data in the Chrome browser [32]. This included data

[1]The privacy settings in some use cases could not be extracted because the supporting papers did not provide enough implementation details.

such as crash reports, homepage settings, and active system processes, supporting research into malware infections by enabling the identification of compromised machines through correlations with users' browsing histories. RAPPOR applied a local DP model by collecting privatized data directly on user devices, thereby ensuring user-level privacy protection. It employed the randomized response technique [38], encoding sensitive values by hashing them into a Bloom filter and introducing randomized noise to preserve DP. Similarly, Microsoft adopted DP in its Windows telemetry data collection service [8], aiming to improve user experience by analyzing patterns in application usage over time (e.g., time spent in specific apps). Equivalent privacy-preserving telemetry mechanism was also developed by Mozilla for the Firefox browser using the Prio framework [58], but its deployment details have not been disclosed. RAPPOR was officially deprecated in 2021 and replaced with more advanced DP-based telemetry solutions.

Extending this trend, Google also integrated DP into its search engine services. To that end, it developed a streaming DP framework that continuously released large-scale differentially private histograms using the Differential Privacy SQL Pipelines (DP-SQLP) algorithm [57], ensuring user-level DP protection. DP-SQLP was deployed to generate two types of statistics. First, Google continuously released product page-view counts (i.e., user impressions), which served as a key signal for *Google Shopping* to prioritize page crawling and update critical information, such as prices and availability, thereby enhancing the shopping experience. Second, DP was applied in *Google Trends* [57] to display trending queries while preserving DP.

Apple used DP for private inference of Safari default auto-play policies for websites that auto-play videos [7], predicting whether to mute or auto-play the sound. Furthermore, Apple used DP to privately identify high-resource-consuming web domains in Safari, i.e., domains that are more likely to create high energy or memory consumption [7]. Apple achieved DP by removing user identifiers and any timestamps, in addition to the use of sketching techniques to perturb values and reduce dimensionality. Apple implemented the DP protection with a privacy unit of a single data collection event (e.g., website visit), providing event-level local DP protection. However, its guarantees can be translated to user-day units because each user device practically sends a limited number of data events per day. In 2023, Apple integrated DP in iOS to identify frequently photographed iconic locations across users, enabling the Photos app to automatically select representative images for features such as Memories, all while protecting individual user data [59]. Apple's learning pipeline ensured event-level DP by processing data on-device, where each device encoded a location-category pair (e.g., a person in New York), applied random noise for local DP, and later generated frequency-based insights for the selection of key photos.

Recently, the Wikimedia Foundation implemented DP to release daily Wikipedia page visits at a country level, based on current and historical Wikipedia page views [60]. This has revolutionized the way the Wikimedia Foundation exposes information to the public about its projects, allowing it to release new statistics on a large scale. According to the Wikimedia Foundation, more than 135M statistics about Wikipedia page visits have been published, aggregating 325B page views in total [60]. This release has been expanded to include differentially private statistics on editor activity (e.g., edit counts) by Wikipedia project and country [61], including a one-time dataset requested by the Russian Wikimedia community to support analysis of the editor landscape in Russia [62].

A similar behavioral analytics use case was announced by LinkedIn, which employs Apache Pinot (a real-time distributed analytics platform) to provide data analysts with interactive access to aggregated insights on user engagement [16]. Through this approach, LinkedIn aims to support marketing analytics applications while ensuring user-level DP protection. To enhance privacy protection and prevent averaging attacks, LinkedIn capped the number of queries an analyst can issue, using a privacy unit defined as one user per analyst per month. This restriction ensures that repeated queries cannot be used to infer individual information through result averaging. In a related deployment, LinkedIn added analytics [63] to each post to analyze user engagement metrics, such as views and shares, while preventing identification of the post's viewers. To that end, LinkedIn integrated DP into their analytics framework so that insights derived from user interactions preserve user confidentiality, thereby upholding privacy standards and user trust. However, the exact DP parameters and privacy units of that analysis were not disclosed. Following a similar strategy, Facebook released a large-scale, differentially private dataset detailing user interactions with over 38 million publicly shared URLs on Facebook between 2017 and 2019 [64]. This differentially private dataset enabled researchers to analyze demographic trends and engagement patterns using aggregated metrics on views and shares.

B. Language & Communication

DP has been increasingly adopted in language and communication technologies to enhance user experiences while safeguarding individual privacy. For example, Apple embedded a local differentially private data collection mechanism in its iOS system to collect statistics about Emoji usage [6], [7] and use it to improve Emoji keyboard ordering (i.e., pushing more popular Emojis to the top of the keyboard layout). In addition, Apple used DP to improve on-device lexicons of previously unknown words typed using QuickType (Apple's predictive keyboard) [7]. These applications employed event-level local DP, where each individual data collection event (e.g., a typing instance) served as the privacy unit.

A related deployment was carried out by Microsoft, who has recently started to apply DP to machine learning prediction tasks. One prominent example is the training of deep learning models on Message-Reply (MR) pairs extracted from emails and chats to suggest reply completions within Microsoft Office tools [65]. The training process preserved user-level DP protection, masking each user's contributions. More recently, Google employed Federated Learning [66] combined with user-level DP for next-word prediction in its Gboard virtual keyboard app [9]. This approach enhanced the typing experience on Android devices through features such as multilingual word suggestions and smart text selection [67].

C. Geo-Location, Mobility & Healthcare

Google leveraged its success with DP implementations to handle location data in a variety of use cases. For instance, DP has been extensively used for releasing mobility reports during the COVID-19 pandemic to aid health researchers in understanding its impact on the population. To that end, Google released community mobility reports about the daily change in mobility patterns of Google users during the COVID-19 pandemic (reflecting work-from-home or stay-at-home policies) [11]. In addition, Google released international urban mobility patterns to study how human movement patterns vary across sociodemographic regions to improve urban sustainability [68], [69]. This study on mobility patterns was expanded into the broader Environmental Insights Explorer (EIE) project, which publicly released global city-level statistics on human mobility and environmental factors, including carbon emissions from buildings and transportation, as well as solar potential [70]. Google enhanced this analysis using Federated Analytics (FA) [30] to process large-scale aggregated location histograms with user-level DP. This was achieved by first locally clipping user location data, then aggregating it on a centralized server with differentially private noise, effectively masking sensitive fine-grained details. In a similar use case, Uber deployed a DP-based system enabling safe queries on customer geospatial data, later formalized as the CHORUS framework [71], [72].

The adoption of DP in healthcare-related applications has progressed more slowly than in other domains. This can be attributed to the inherent sensitivity of medical data and the challenges healthcare organizations face in implementing privacy-preserving techniques that comply with legal frameworks such as the Health Insurance Portability and Accountability Act (HIPAA) [73]. Despite the legal and technical challenges, Apple integrated DP mechanisms into its HealthKit app to enhance its functionality [7]. Specifically, Apple employed local DP to count the most common health data types that users were monitoring over time while ensuring that no sensitive information could be learned about users' medical conditions. For example, Apple reported sleep analysis, heart rate, and calories burned as the most monitored user indicators.

Spectus utilized DP to protect the statistics concerning evacuation rates of people during natural disasters [74], such as Hurricane Irma, which hit the U.S. East Coast in 2017. By measuring the percentage of evacuated residents, common destinations, and traveled distance, Spectus provided emergency services with insights into the effect of natural disasters on the population to improve their response in case of an emergency. Another notable example in the healthcare domain comes from Google, which employed differential privacy during the COVID-19 pandemic to help health organizations better understand the needs of their communities. Specifically, Google utilized its search engine to publish daily count trends of Google searches pertaining to COVID-19 medical symptoms across different geographical areas [75]. Moreover, in an effort to raise awareness of the necessity of vaccines and establish a status for COVID-19 vaccination rates, Google issued differentially private daily count trends of searches concerning COVID-19 vaccinations, vaccination intent, and related medical side effects [10]. Furthermore, a joint effort by Google and Apple produced the Exposure Notification Privacy-preserving Analytics (ENPA) system [76] that enabled notifications of exposure to positively diagnosed COVID-19 patients without disclosing any personally identifiable information. Google and Apple's notification model was protected with shuffle DP [35], incorporating both local randomized response and central aggregation.

D. Census, Demographics & Economy

Governmental organizations, such as the U.S. Census Bureau, have also been utilizing DP for statistical data analysis. The U.S. Census Bureau was a pioneer in deploying DP for governmental applications, unprecedentedly developing "On-TheMap" [77]—an interactive mapping system for querying residence and workplace patterns of the U.S. population. The system provided demographic and employment data of residents within a selected area, including occupations and salaries, which could be queried by race, ethnicity, education, and gender. "OnTheMap" presented the privatized statistics based on differentially private synthetic data, which was generated based on real census data to preserve its statistical properties without risking the data subjects' privacy. Following this successful application, the U.S. Census Bureau developed a Disclosure Avoidance System (DAS) [13], offering centralized DP, to guard against sensitive information disclosure in the summary data of the 2020 Decennial Census. In one of its deployments, the U.S. Census Bureau released statistics about the employment of post-secondary education graduates [78], which was released based on obtained data from the Longitudinal Employer-Household Dynamics (LEHD) dataset (quarterly earnings records from 50 states), the Census's Quarterly Workforce Indicators, and graduate records from education partners. Specifically, count and percentiles of earnings per the combination of degree level, degree field, institution, and graduation year were privatized and published. Moreover, the U.S. Census Bureau released demographic data in its 2020 Census Redistricting Dataset [79] and the Demographic and Housing Characteristics File (DHC), which provides information on population and household characteristics [12]. A separate differentially private version of the DHC was published for person- and housing unit-level data, offering DP at the citizen-level.

The U.S. Census Bureau has recently announced a novel use case in which it has used disclosure avoidance mechanisms to produce demonstration tables for the County Business Patterns (CBP) data product [80]. These tables aggregated vital economic statistics of business establishments in the U.S., including establishment counts, annual payroll, and employment size. Due to the skewness of the data and its heavy-tailed distribution, per-record DP was employed, according to which a "sliding establishment protection" was provided with varying privacy guarantees for each establishment. In a related effort, the U.S. Internal Revenue Service (IRS) employed DP to release college graduate income statistics through the U.S. College Scorecard [81], though specific implementation details

were not publicly disclosed. This deployment was carried out using the Tumult Analytics platform [82].

In light of the U.S. Census Bureau's extensive use of differentially private mechanisms, other bureau offices have started investigating DP to protect their releases. The U.K. Office for National Statistics (ONS) initiated a pilot study for examining the potential of DP for mortality data statistics as part of the 2021 UK census [83]. The Statistics Bureau of Japan published similar intents [14], examining the utility of DP methods for official Japanese statistical data, including geographical data from the Japanese Population Census. Although the Statistics Bureau of Japan did not specify the final DP parameters used in its production implementation, experiments with a wide range of ε values were conducted, spanning from 0.1 to 100 [15]. The Australian Bureau of Statistics (ABS) has implemented a customized differentially private perturbation mechanism to the ABS TableBuilder [84], an interactive analytic tool for the generation of count tables based on census data. DP protection was only implemented for a single TableBuilder counting query, but the protection of queries under a dynamic environment is currently under further research. Recently, Israel's Ministry of Health released a differentially private synthetic dataset based on the 2014 National Registry of Live Births [85], safeguarding the identities of mothers and newborns at a birth event-level.

Recent commercial deployments of differential privacy by various companies in the economic domain have also begun to emerge. In the smart energy sector, the Energy Differential Privacy (EDC) project by Recurve-OhmConnect [18] released residential energy consumption statistics derived from smart meter data. Specifically, differentially private estimates of average energy load and percent load change were computed using data from 4,948 non-solar electric meters, with a privacy unit of a user (meter) per day. In the labor market domain, LinkedIn applied DP to publish monthly hiring statistics segmented by industry and region, supporting labor market analysis during the COVID-19 pandemic [86]. Additionally, LinkedIn released differentially private data on in-demand skills for top trending jobs to assist job seekers in career planning.

In 2024, LinkedIn introduced a privacy-preserving approach using Randomized Response to preserve local DP while releasing U.S. race and ethnicity data [87]. This initiative aims to enhance AI fairness by enabling comparisons of system performance across demographic groups, particularly when race and ethnicity data are limited. To achieve this, LinkedIn developed the Privacy-Preserving Probabilistic Race/Ethnicity Estimation (PPRE), which combines the Bayesian Improved Surname Geocoding (BISG) model with self-reported demographic data. DP was enforced by applying randomized response to users' self-reported race values, creating a privacy-preserving dataset of race probabilities.

VI. A Comparison of DP Configurations: Academic Research vs. Industry

A. High-Level Comparison Between DP Configurations

While academic research and commercial applications of DP share common algorithmic foundations, their final configurations often diverge in practice (Figure 1). We can deduce that academic studies often adopt a more conservative range of ε values, in contrast to the wider variability seen in industrial implementations. This divergence in the range of chosen ε values is evident across both complex analytical tasks and traditional descriptive statistical analyses, such as histogram computations. For instance, academic studies have commonly used small privacy budgets ranging from 0.05 to 0.2 for differentially private histogram releases (i.e., counting queries) [88]. In contrast, similar types of histogram analyses conducted by industry actors such as Google [10], [11], [75] and Apple [7] have used significantly higher privacy budgets, with values reaching up to 8 in the case of Apple's website autoplay count release [7]. Similar discrepancies appear in smart energy applications: while empirical research has configured DP with ε values ranging from 0.1 to 2 [89], [90], industrial deployments such as the Recurve project have used substantially higher budgets, with $\varepsilon = 6.8$ [18] per day, thereby offering weaker formal privacy guarantees.

Examining the DP configuration trends (in terms of ε) over the last 5 years (Figure 1), we identified several noteworthy patterns. First, academic research on differential privacy has consistently adopted strong privacy guarantees, with privacy loss budgets commonly in the range $0.05 \leq \varepsilon \leq 5$ with a median of $\varepsilon = 0.1$ (S.D.$= 2.5$). Second, a notable divergence is observed in the configurations employed by commercial and governmental/non-profit organizations. Commercial organizations have deployed ε values ranging from 0.5 to 10 (median is $\varepsilon = 2.7$; S.D.$= 7.5$), with most configurations offering DP guarantees over a temporal scope (e.g., per day or per month). In contrast, government agencies and non-profits have used significantly larger ε values for large-scale statistical releases (median is $\varepsilon = 5.8$; S.D.$= 13.7$), with the U.S. Census Bureau allocating a notably high $\varepsilon = 35.62$ for the County Business Patterns (CBP) dataset [80]. Moreover, both sectors exhibit a trend of increasing ε values over time, aiming to support more complex or frequent data analyses by relaxing privacy constraints. As organizations in the future will report more deployments, a more accurate trend could be obtained over time, particularly among governmental agencies, where adoption of DP for new census analysis may be slower than in commercial companies.

B. Reasoning Behind DP Configuration Differences

We analyzed the surveyed papers and industry reports to identify the reasons behind selected DP configurations (primarily ε), that can explain the distinct configuration choices observed in academic research and industrial deployments (Table III). While the majority of our surveyed academic papers (59%) and a large portion of commercial deployments (25%) did not provide any justification for their choice of DP parameters, the rest deliberated on their selection choices. In academic research, ε values were often chosen arbitrarily or guided by best practices from prior literature, e.g., stating that "typically, $\varepsilon \leq 0.1$ is considered strong and $\varepsilon \geq 10$ is considered weak" [94]. Only 6% of academic papers provided goal-oriented rationales, selecting parameters to meet predefined utility or privacy requirements. Most of

TABLE III
DISTRIBUTION OF JUSTIFICATION TYPES FOR DP CONFIGURATION SELECTION IN ACADEMIC RESEARCH AND COMMERCIAL (COMM.) OR GOVERNMENTAL/NON-PROFITS (GOVT./NP) PRACTICE. PERCENTAGES (IN PARENTHESES) INDICATE PROPORTION OF COUNTS FROM TOTAL.

Justification type	**Examples**	**Comm.** ($n = 32$)	**Govt./NP** ($n = 10$)	**Academic** ($n = 81$)
No justification	–"Without explanation, we set $w = 200$ and $\varepsilon = 1$ for all experiments." [91]	8 (25%)	2 (20%)	48 (59%)
Arbitrary (unsubstantiated reasoning)	–"From the definition a choice of $\varepsilon \leq 1$ seems reasonable." [92] –"ε is uniformly randomly drawn from $E_1 = \{0.25, 0.5, 0.75\}$...." [93]			13 (16%)
Community best practice	–"Typically, $\varepsilon \leq 0.1$ is considered strong and $\varepsilon \geq 10$ is considered weak." [94] –"Considering past real-world differentially private releases as a benchmark, we aimed to a total privacy loss budget of $\varepsilon < 10$..." [85]	6 (19%)	1 (10%)	12 (15%)
Privacy/utility trade-off tuning	–"The Vaccination Search Insights are designed to maintain the privacy of our users while releasing [...] data that is as accurate and useful as possible." [10] –"We managed to obtain a significantly better model, while ensuring that users' data stays private." [67]	9 (29%)		3 (4%)
Predefined privacy/utility constraints	–"[...] ε was determined stochastically to achieve [...] 1% error..." [18] –"We take a utility-first approach, as the end application requires an average relative weighted error of $\approx$ 3% to be useful" [70]	5 (15%)	4 (40%)	5 (6%)
Policy, regulatory, or ethical compliance	–"[...] we adhere to strict policy regarding the privacy budget." [95] –"The TDA parameters for the published [...] data were primarily policy-driven. In setting these parameters, the agency had to consider and balance its countervailing obligations to produce high-quality statistics while also protecting the confidentiality of census respondents..." [12]	4 (12%)	3 (30%)	

these cases were observed in specific empirical applications, such as smart energy metering, where utility is critical due to the risks of inaccurate consumption data, which can result in inflated bills (e.g., "[...] using a utility requirement of 5%, the achieved privacy level is..." [96]). This type of justification was also observed among 40% of the surveyed deployments by governmental and non-profit organizations.

In contrast to academic research, 29% of the surveyed industrial deployments provided more concrete justifications, often grounded in empirical analysis of utility versus privacy objectives. This reflects practical considerations in which DP configurations are shaped by privacy-utility trade-offs [20], where parameters are selected to balance analytical utility and privacy risk. Hence, organizations tend to orient their choices of ε around one of three strategies: utility-focused, privacy-focused, or a balanced approach. For instance, the Wikimedia Foundation adopted a utility-oriented approach, selecting parameters to "optimize the global utility metrics" [60]. The U.S. Census Bureau initially emphasized privacy in the release of employment statistics, requiring that data "not include personally identifiable information" [78], but later transitioned to a more utility-focused approach in its 2020 redistricting data release, developing the Disclosure Avoidance System (version 12.2) which "represents a relatively high privacy loss [...] at the expense of greater privacy loss..." [79]. Other commercial organizations, such as Recurve, prioritized privacy, "erring on the side of caution" [18] when setting ε. In contrast, LinkedIn presented a more balanced approach, noting that their differentially private algorithms "have better accuracy and privacy tradeoffs" [63].

Moreover, 12% of surveyed commercial deployments and 30% of the governmental deployments described a compliance with policy, regulation, or ethical guidelines as their primary motivation for the selected DP configuration. These policies may be internal to the company or external regulations with which organizations must comply (such as GDPR or the United States Code). For example, Microsoft's Privacy-Preserving Machine Learning (PPML) policy limits privacy loss to $\varepsilon = 4$ over six months for any contributing user [95]. Similarly, as a governmental body that is subject to the public transparency, the U.S. Census Bureau similarly justified its Post-Secondary Employment Statistics release by stating that "In carrying out the public reporting and disclosure requirements of this Act (Title 13 of the U.S. code), the Commissioner shall use appropriate statistical disclosure limitation techniques necessary to ensure that the data released to the public cannot include personally identifiable information...." [78]. A similar proportion (19%) of deployments cited community best-practice approaches informed by prior deployments. For example, Apple provided a partial justification for all of its deployments stating that "Our choice of ε [...] these values are consistent with the parameters proposed in the [...] research community" [7]. Similarly, the Israeli Ministry of Health justified its configuration for live births data using precedents from prior industry releases: "Considering past real-world differentially private releases as a benchmark [53], we aimed to a total privacy loss budget..." [85]).

The observed rationales suggest several possible explanations for the differing choices of ε values. First, community best practices shape parameter choices differently in commercial, governmental or academic settings. These differences results in distinct configuration norms. Academic research on DP initially focused on theoretical work, where the choice of ε was often arbitrary or guided by research community practices rather than rigorous justification, motivating their choice of smaller εs. It later expanded to applied studies that adopted empirical methodologies and used a wider range of ε values, especially in recent years (Figure 1). In contrast, industry organizations used prior deployments or policies from similar organizations as benchmarks, such as Microsoft's PPML policy [95], which adopts a higher privacy budget of $\varepsilon = 4$. Second, in contrast to academic research which deals with

relatively small and static datasets, commercial deployments tend to process much larger data volumes continuously over time (e.g., telemetry analysis by Microsoft [8] collects data daily). Although privacy loss is often capped per time unit (e.g., per user per day), the cumulative effect leads to higher overall budgets (weaker privacy guarantees) than in non-temporal use cases. Moreover, many commercial deployments adopt local DP, with user-specific ε values assigned by the data processor rather than the user. This contrasts with academic research, which typically employs central DP setups with markedly different ε choices. Specifically, deployed ε values were significantly higher in local DP deployments (12 cases) than in central DP deployments (38 cases; $U = 306.5$, $p < .05$). This observation is consistent with previous literature [23] that suggests that local models require higher εs to mitigate the greater noise added individually by users.

C. Examples of DP Configuration Shifts from Research to Commercial Deployment

Several cases exist where the same entity contributed to both an academic publication and a commercial deployment, offering an opportunity to demonstrate the differences between academic research and commercial configurations. These differences may be attributed to the need of commercial organizations to adapt DP configurations when transitioning from research to practice, in response to operational demands, regulatory constraints, and internal policies that directly influence configuration choices [97]. For example, Microsoft's research team developed the Differentially Private Set Union (DPSU) algorithm that constructs a large subset of the union of user-contributed sets, enabling efficient selection of frequent items under user-level DP. The DPSU algorithm was employed to construct differentially private n-gram histograms from user-generated text, enabling the identification of frequent phrases to train a reply suggestion model for Microsoft Office services (i.e., emails and messages) [65]. The algorithm was evaluated under a user-level (ε, δ)-DP guarantee using ε values ranging from 0.5 to 4 with $\varepsilon = 3$ and $\delta = 10^{-10}$ as the representative setting. However, the deployed configuration by the same research team at Microsoft used $\varepsilon = 4$, representing the upper bound of the tested range and the maximum allowed budget under Microsoft's Privacy-Preserving Machine Learning policy [95]. In addition, δ was relaxed to 10^{-7}. In this case, moving from research to commercial deployment involved increasing the privacy budget from $\varepsilon = 3$ in research experiments to $\varepsilon = 4$ in the deployed product, resulting in a degradation of the privacy guarantees by 33%.

Google's deployment of DP in the Google Trends tool using the DP-SQLP streaming framework [57] illustrates a distinct narrative within deployment practices. Initially, the release of private histograms from simulated user activity using DP-SQLP was evaluated under a user-level ($\varepsilon = 6$, $\delta = 10^{-9}$)-DP. However, during deployment of that framework for the Google Trends tool, the configuration was tightened: each query was processed under $\varepsilon = 2$, $\delta = 10^{-10}$, and users were limited to one contribution per query. An additional pre-threshold of 50 unique users was applied before noise was added, excluding low-frequency queries, thereby providing stronger privacy guarantees than in the experimental setup. This choice of a lower ε in deployment was driven by practical privacy constraints for user-level protection in a production environment, which involves a continuous stream of sensitive and diverse real-user data (e.g., search queries related to medical conditions). Since each query incurs $\varepsilon = 2$, repeated user contributions lead to accumulating privacy loss via composition. This necessitates a tighter control over the privacy loss budget, particularly for low-frequency Google Trend queries that pose greater privacy risks to users.

VII. Discussion

As no "silver bullet" guidelines exist for selecting ε, this survey highlights the diversity of DP configurations in practice and examines the underlying rationales. This section discusses key findings and proposes directions for future research.

A. Key Limitations and Gaps in Commercial and Governmental DP Configurations

1) Limited Transparency of DP Configuration Details: Due to the lack of a unified knowledge base for DP implementations [19], we manually extracted configuration details from various sources, revealing transparency issues in reporting of privacy parameters, such as ε and δ. First, as our analysis of practical justifications for privacy parameter selection revealed, 59% of academic papers and 25% of commercial deployments provide no justification for their DP configurations. Second, the absence of standardized terminology in reporting DP configurations and implementation details hindered the precise extraction of DP parameters. Specifically, details of the employed privacy protection unit were partially deduced from the text and were post-processed to measure protection in similar units to those used in other implementations. For example, each count query to LinkedIn's labor marketing system [86] covered three months of data, but statistics were published monthly, and hence the privacy guarantees were reported for a single month. Second, a difference in reporting of (ε, δ) values was observed for use cases implementing different DP variants. For example, in zero-concentrated DP (zCDP), guarantees are expressed using the ρ parameter but can be converted to the standard (ε, δ)-DP parameters (according to [44]). This may create a gap in perception of the provided privacy guarantees since the scale of parameter values can vary across DP definitions, e.g., $\rho = 0.48$-zCDP with $\delta = 10^{-10}$ is translated to $(\varepsilon = 6.83, \delta = 10^{-10})$-DP. Apart from difficulty in reproducibility, the lack in transparency can also have an effect on the reputation of organizations and create incredulity, which may reduce the willingness of potential data subjects to share personal data.

To mitigate the aforementioned transparency issues in the context of DP, privacy "Nutrition Labels" [98] can be used, similar to "fact box" visualizations in healthcare that are often used to simplify complex tabular data. Originally proposed for privacy policy communication, these labels provide clear privacy guarantees and facilitate easier access to policy information for data consumers. Moreover, data practitioners implementing DP can use such labels to compare DP parameters

between implementations, specifically the parameters of ε and δ. A possible direction for future work is to expand privacy labels to convey more detailed DP guarantees and tailor them for data practitioners with varying levels of DP expertise.

2) No User-Oriented DP Configurations: We found no evidence of DP implementations where data subjects shared their personal privacy preferences with data processors. As a result, current DP configurations reflect only organizational definitions of privacy, which may not align with individuals' expectations. When data provision is mandatory, such as in national censuses, individuals lack control over its use, leading to a mismatch in privacy perception. In contrast, opt-in models enable user-defined privacy preferences. For example, Apple's local DP implementation [7] could, in principle, let users control the ε value for noise addition. However, identifying relevant preferences and integrating them into differential privacy remains a key challenge and avenue for future research.

B. ε as a Hyper-Parameter or a Deliberate Design Choice

Our analysis highlights distinct approaches to selecting ε and δ in academic versus commercial or governmental settings, reflecting differing priorities and methodological paradigms. Based on our surveyed academic research papers, we conclude that the majority of research treats ε (and also δ) as a hyper-parameter, i.e., an external configuration variable which is not learned from the data or model itself and is set prior to the analysis process. This approach can be problematic, as the final configuration may be biased by the initial selection, potentially unsupported by the data or task requirements. Such biases can carry over into real-world deployments when academic research is applied in practice. Therefore, this underscores the need for a context-aware approach when selecting DP parameters, as aligning privacy budgets with social norms can lead to more appropriate configurations [99]. For example, the same privacy configuration may be appropriate in a commercial app but unacceptable in a government census.

In contrast to academic work, commercial and governmental deployments treat DP parameters as deliberate design choices, integrating them into the core system architecture. In such deployments, ε, δ, and the noise mechanism were selected based on task-specific needs, requiring organizations to carefully balance privacy and utility. Several deployments in our survey explicitly acknowledged this trade-off [20] in their parameter choices. For example, LinkedIn applied a privacy-by-design approach in its race and ethnicity estimation system, stating that "the measurement test must have privacy by design at its core" [87], selecting $\varepsilon = 4.5$ to ensure that "information must be comprehensive and useful enough to enable equal treatment measurements with respect to race and ethnicity at the aggregate level" [87]. This highlights the need for increased transparency and a structured framework for DP configuration, as proposed by Dwork et al. [19], who advocate for an Epsilon Registry to document implementation choices and guide practitioner decision-making.

C. Standardization of the DP Configuration Process

As indicated by a recent study by Sarathy et al. [100], organizations frequently turn to trial-and-error approaches in selecting DP parameters, highlighting the ongoing challenges in establishing principled DP configurations. Given these challenges, a structured approach to configuring DP may be beneficial, drawing on similar frameworks from machine learning [101] and cybersecurity [102]. Just as ML models undergo testing and refinement to improve robustness against adversarial attacks, DP configurations might benefit from similar iterative adjustments to strengthen protection against inference risks while maintaining utility. A potential workflow for configuring DP could involve five key steps: *Defining the target data analysis*, *Specifying privacy guarantees*, *DP operationalization*, *Evaluation*, and *Documentation*.

This process can begin with defining the data analysis in terms of expected outputs and duration, which influences privacy budget allocation. Consequently, the data processor can define the privacy guarantees, taking into account factors such as potential privacy risks, the protected privacy unit, and the DP mechanism applied. This step helps clarify the level of privacy protection provided. The operationalization phase would then involve setting DP parameters: privacy budget (ε), failure probability (δ in (ε, δ)-DP), and other loss parameters (such as ρ in ρ-zCDP). Once these parameters are configured, evaluation becomes important to ensure that the chosen approach aligns with both data subjects' privacy expectations and organizational policies. This evaluation typically considers the privacy-utility trade-off, measuring privacy protection (e.g., probability of privacy breaches) relative to predefined utility metrics (e.g., accuracy). If necessary, adjustments can be made iteratively to refine the balance between these two factors. Finally, documenting the configuration could serve as a reference for future analyses, supporting the broader idea of an Epsilon Registry [19] for benchmarking DP implementations.

VIII. Conclusions

Differential Privacy (DP) has seen growing adoption in both academic research and industry, supporting large-scale data analysis while providing strong privacy guarantees for involved users. This survey presented a systematic literature review of recent deployments of DP for applications by commercial and governmental organizations, and compared their DP parameters with corresponding academic research. Our analysis reveals that National Statistical Offices, such as the U.S. Census Bureau, have used a wider range of privacy budget values, reflecting growing demand for public data releases and more flexible DP configurations. On the other hand, commercial companies reflected a different privacy-utility orientation in their DP configurations, allocating ε values over a narrower range for different use cases. In contrast to commercial or governmental deployments, academic research on DP has consistently used lower ε values, often following community best practices that recommend $\varepsilon \leq 1$. By expanding the current knowledge on DP applications and their challenges, our analysis supports data practitioners and policymakers in benchmarking DP configurations and advancing the vision of a unified Epsilon Registry for practical DP configurations.

TABLE IV

REAL-WORLD IMPLEMENTATIONS OF DIFFERENTIAL PRIVACY (DP) DEPLOYED IN PRODUCTION ENVIRONMENTS BY COMMERCIAL, GOVERNMENTAL, AND NON-PROFIT ORGANIZATIONS. DP IS FORMALLY DEFINED AS EITHER PURE ε-DP (A), APPROX. (ε, δ)-DP (B) OR ZCDP (C). PERTURBATION IS ACHIEVED USING EITHER THE LAPLACE MECHANISM (①), GAUSSIAN MECHANISM (②), GEOMETRIC MECHANISM (③), RANDOMIZED RESPONSE/ENCODING (④), EXPONENTIAL MECHANISM (⑤) OR OTHER (⑥). LEGEND: DISC.= DISCRETE; CONT.= CONTINUOUS; ○ = NO; ● = YES.

Use Case					DP Configuration											
Organization	Year	Use Case Description	Domain	Analysis type	Granularity		Privacy unit	Mode			DP variant	Perturb. Mechanism			ρ-zCDP Param.	(ε, δ)-DP Params.	Other details
					User	Event		Central	Local	Hybrid		Mech.	Disc.	Cont.			
Wikimedia Foundation	2023	Wikipedia page visits [60]	Web & Browsing Behavioral Analytics	Descriptive Analytics	●	○	User per day	●	○	○	Ⓒ	②	●	○	0.015	$(0.998, 10^{-7})$	Current page views
					●	○	30 page views per day	●	○	○	Ⓐ	③	●	○		$(1, 0)$	Historical 2017-2023
					●	○	300 page views per day	●	○	○	Ⓐ	③	●	○		$(1, 0)$	Historical 2015-2017
	2023	Wikipedia editor activity statistics (from all countries) [61]	Web & Browsing Behavioral Analytics	Descriptive Analytics	●	○	Editor (user) of project in country per month/week	●	○	○	Ⓐ	①	○	●		$(2, 0)$	All countries
	2023	Wikipedia editor activity statistics (from Russia) [62]	Web & Browsing Behavioral Analytics	Descriptive Analytics	●	○	User account per month	●	○	○	Ⓐ	①	○	●		$(0.1, 0)$	Russia (one-time)
Apple	2023	Iconic scenes and photo selections for iOS apps [59]	Web & Browsing Behavioral Analytics	Behavioral Analysis	○	●	—	○	●	○	Ⓐ	④	●	○		$(1, 1.5 \times 10^{-7})$	
	2017	Emoji suggestions [6], [7]	Language & Communication	Predictive Modeling	○	●	Event (typing) per day	○	●	○	Ⓐ	④	●	○		$(4, 0)$	Max. 1 event per day
	2017	QuickType suggestions [6], [7]	Language & Communication	Predictive Modeling	○	●	Event (typing) per day	○	●	○	Ⓐ	④	●	○		$(8, 0)$	Max. 2 events per day
	2017	Lookup hints from iOS search suggestions [6], [7]	Language & Communication	Predictive Modeling	○	●	Event (typing) per day	○	●	○	Ⓐ	④	●	○		$(4, 0)$	Max. 2 events per day
	2017	Safari Autoplay intent detection [6], [7]	Web & Browsing Behavioral Analytics	Behavioral Analysis	○	●	Event (website) per day	○	●	○	Ⓐ	④	●	○		$(8, 0)$	Max. 2 events per day
	2017	Safari energy consumption [6], [7]	Web & Browsing Behavioral Analytics	Descriptive Analytics	○	●	Event (website) per day	○	●	○	Ⓐ	④	●	○		$(4, 0)$	Max. 2 events per day
	2017	Health type usage in HealthKit app [6], [7]	Geo-Location, Mobility & Healthcare	Descriptive Analytics	○	●	Event (measure) per day	○	●	○	Ⓐ	④	●	○		$(2, 0)$	Max. 1 event per day
Google	2024	Environmental Insights Explorer [70]	Geo-Location, Mobility & Healthcare	Descriptive Analytics	●	○	User per week	○	○	●	Ⓐ	①	○	●		$(2, 0)$	
	2024	Gboard Out-Of-Vocabulary word discovery [103]–[105]	Language & Communication	Predictive Modeling	○	●	Single word (Max. 60 words in 60 days)	○	●	○	Ⓐ	⑤	●	○		$(10, 0)$	Local noisy user votes
					●	○	○	Ⓑ	①	○	●					$(0.315, 10^{-10})$	Noisy OOV counts (en-US)
	2024	User impressions for Google Shopping [57]	Web & Browsing Behavioral Analytics	Descriptive Analytics	●	○	User per day	●	○	○	Ⓑ	②	●	○		$(1, 10^{-9})$	
	2024	Selection of queries for Google Trends [57]	Web & Browsing Behavioral Analytics	Descriptive Analytics	●	○	User-query	●	○	○	Ⓑ	②	●	○		$(2, 10^{-10})$	
	2023	Smart text selection for Android [67]	Language & Communication	Predictive Modeling	●	○	—	○	○	●	—	—	—	—		—	
	2023	Next-word Prediction (and Smart Compose) in Gboard [56], [106], [107]	Language & Communication	Predictive Modeling	●	○	Device per 24 hours	●	○	○	Ⓒ	②	●	○	0.0143	$(0.994, 10^{-10})$	Portuguese (Brazil)
											Ⓒ	②	●	○	0.0143	$(0.994, 10^{-10})$	Spanish (Latin America)
											Ⓒ	②	●	○	0.150	$(3.424, 10^{-10})$	Spanish (Spain)
											Ⓒ	②	●	○	0.250	$(4.799, 10^{-10})$	English (US)
											Ⓒ	②	●	○	0.480	$(6.827, 10^{-10})$	German (Germany)
	2021	COVID-19 vaccination search insights [10]	Geo-Location, Mobility & Healthcare	Descriptive Analytics	●	○	User per day	●	○	○	Ⓑ	②	○	●		$(2.19, 10^{-5})$	
	2020	COVID-19 community mobility reports [11]	Geo-Location, Mobility & Healthcare	Descriptive Analytics	●	○	User per day	●	○	○	Ⓐ	①	○	●		$(1.76, 0)$	Public places statistics
											Ⓐ	①	○	●		$(0.44, 0)$ } $= (2.64, 0)$	Residential places statistics
											Ⓐ	①	○	●		$(0.44, 0)$	Workplace statistics
	2020	COVID-19 search trends symptoms dataset [75]	Geo-Location, Mobility & Healthcare	Descriptive Analytics	●	○	User per day	●	○	○	Ⓐ	①	○	●		$(1.68, 0)$	
	2019	Urban mobility patterns [68], [69]	Geo-Location, Mobility & Healthcare	Behavioral Analysis	●	○	Whether user made a trip per week	●	○	○	Ⓑ	①	○	●		$(0.66, 2.1 \times 10^{-29})$	Max. 1 trip within a week for a given user
	2014	Google Chrome browsing statistics (RAPPOR) [32]	Web & Browsing Behavioral Analytics	Descriptive Analytics	●	○	User per day	○	●	○	Ⓐ	④	●	○		$(25.63, 0)$	30 min reporting intervals
Google and Apple collab.	2021	COVID-19 exposure notification privacy-preserving analytics [76]	Geo-Location, Mobility & Healthcare	Behavioral Analysis	●	○	User per day	○	●	○	Ⓐ	④	●	○		$(8, 0)$	
Mozilla Firefox	2018	Firefox browser data telemetry with Prio [54], [58]	Web & Browsing Behavioral Analytics	Descriptive Analytics	●	○	—	●	○	○	—	—	—	—		—	
Facebook	2020	Facebook user interactions with web pages (Full URLs dataset) [64]	Web & Browsing Behavioral Analytics	Descriptive Analytics	●	○	User	●	○	○	Ⓒ	②	○	●	0.073	$(1.453, 10^{-5})$	
	2020	Movement Range Maps [108]	Geo-Location, Mobility & Healthcare	Descriptive Analytics	●	○	User per day	●	○	○	Ⓐ	①	○	●		$(1, 0)$ } $= (2, 0)$	Change in movement metric
								●	○	○	Ⓐ	①	○	●		$(1, 0)$	Stay-put metric
Microsoft	2023	The global victim-perpetrator synthetic dataset [109], [110]	Census, Demographics & Economy	Synthetic Data	●	○	User (victim)	●	○	○	Ⓑ	②	○	●		$(12, 5.8 \times 10^{-6})$	
	2021	U.S. Broadband coverage [17]	Web & Browsing Behavioral Analytics	Descriptive Analytics	●	○	User	●	○	○	Ⓐ	①	○	●		$(0.1, 0)$ } $= (0.2, 0)$	Speed telemetry query
								●	○	○	Ⓐ	①	○	●		$(0.1, 0)$	Services devices query
	2020	Reply AI suggestion [65]	Language & Communication	Predictive Modeling	●	○	User per 6 months	●	○	○	Ⓑ	②	○	●		$(4, 10^{-7})$	
	2017	Windows telemetry [8]	Web & Browsing Behavioral Analytics	Descriptive Analytics	●	○	User per day	○	●	○	Ⓐ	④	●	○		$(1.672, 0)$	6 hour reporting
LinkedIn	2024	Privacy-preserving race and ethnicity estimation [87]	Census, Demographics & Economy	Descriptive Analytics	●	○	User	○	●	○	Ⓐ	④	●	○		$(4.5, 0)$	
	2023	Privacy-enhanced post analytics [63]	Census, Demographics & Economy	Descriptive Analytics	●	○	—	●	○	○	Ⓒ	②	○	●		—	
	2020	LinkedIn's audience engagements API [16]	Web & Browsing Behavioral Analytics	Behavioral Analysis	●	○	User	●	○	○	Ⓑ	① ⑤	○	●		$(0.15, 10^{-10})$	Single query
					●	○	User per month per analyst	●	○	○	Ⓑ	① ⑤	○	●		$(34.90, 7 \times 10^{-9})$	Multiple queries (monthly)
	2020	LinkedIn's labor market insights [86]	Census, Demographics & Economy	Descriptive Analytics	○	●	Event per month	●	○	○	Ⓑ	①	○	●		$(4.8, 4 \times 10^{-10})$	Top 20 employer metric
											Ⓑ	①	○	●		$(4.8, 4 \times 10^{-10})$	Top 20 jobs metric
											Ⓑ	⑤	○	●		$(0.1, 10^{-10})$	Top 20 trending skills metric
Recurve	2020	Energy measurement for the OhmConnect virtual power plant [18]	Census, Demographics & Economy	Behavioral Analysis	●	○	User (meter) per day	●	○	○	Ⓑ	②	○	●		$(4, 4.08 \times 10^{-8})$	Average load shape
											Ⓐ	①	○	●		$(0.3, 0)$ } $= (6.8, 4.08 \times 10^{-8})$	SVT for clamping bound
											Ⓐ	①	○	●		$(2.5, 0)$	Percent load change
The US Census Bureau	2023	County Business Patterns (CBP) data product [80]	Census, Demographics & Economy	Descriptive Analytics	●	○	Geo-cell of establishments	●	○	○	Ⓒ	②	●	○	12.06	$(35.62, 10^{-5})$	
	2023	2020 U.S. Demographic and Housing Characteristics File (DHC) [12]	Census, Demographics & Economy	Descriptive Analytics	●	○	Person	●	○	○	Ⓒ	②	●	○	4.96	$(26.34, 10^{-10})$	Person tables (DHCP)
											Ⓒ	②	●	○	7.70	$(34.33, 10^{-10})$	Housing units tables (DHCH)
	2021	2020 census redistricting data [79]	Census, Demographics & Economy	Descriptive Analytics	●	○	Person	●	○	○	Ⓒ	②	●	○	2.56	$(17.14, 10^{-10})$	Person Tables
											Ⓒ	②	●	○	0.07	$(2.47, 10^{-10})$	Housing unit Tables
	2019	Post-Secondary Employment Outcomes (PSEO) [78], [111]	Census, Demographics & Economy	Descriptive Analytics	●	○	Person	●	○	○	Ⓐ	③	●	○		$(1.5, 0)$ } $= (3, 0)$	Graduate earnings percentiles
											Ⓐ	③	●	○		$(1.5, 0)$	Employment flows counts
	2008	OnTheMap [77], [112]	Census, Demographics & Economy	Descriptive Analytics	—	—	Geographical block	●	○	○	Ⓑ	⑥	●	○		$(8.99, 10^{-6})$	Bayesian Sampling
Australian Bureau of Statistics	2019	Single counting query using the ABS TableBuilder [84]	Census, Demographics & Economy	Descriptive Analytics	●	○	Person	●	○	○	Ⓑ	—	●	○		$(1.253, 0.53)$	Cell counts < 7
																$(0.693, 0.04)$	Cell counts ≥ 7
Israel Ministry of Health	2024	Live births synthetic dataset [85]	Census, Demographics & Economy	Synthetic Data	○	●	Single singleton birth	●	○	○	Ⓐ	⑥	●	○		$(4, 0)$ } $\times 2 = (9.98, 0)$	Generative model training with Bayesian networks
											Ⓐ	①	○	●		$(0.99, 0)$	Acceptance criteria
Spectus (by Cuebiq)	2022	Hurricane Irma evacuation rate dashboard [74], [113]	Geo-Location, Mobility & Healthcare	Descriptive Analytics	—	—	—	●	○	○	Ⓐ	①	○	●		$(4, 0)$	Count of total users
											Ⓐ	①	○	●		$(2, 0)$ } $= (10, 0)$	Count of evacuees
											Ⓐ	①	○	●		$(1, 0)$	Count of evacuees per dest.
											Ⓐ	①	○	●		$(3, 0)$	Sum of distances from home
Uber	2017	Internal data analytics (e.g., fraud detection) [71], [114]	Web & Browsing Behavioral Analytics	Descriptive Analytics	—	—	—	●	○	○	Ⓑ	①	—	—		—	Parameters vary by query

This article has been accepted for publication in IEEE Transactions on Knowledge and Data Engineering. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TKDE.2025.3603731

REFERENCES

[1] C. Dwork, F. McSherry, K. Nissim, and A. Smith, "Calibrating noise to sensitivity in private data analysis," in *Theory of Cryptography*, S. Halevi and T. Rabin, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2006, pp. 265–284.

[2] W. Liu, Y. Zhang, H. Yang, and Q. Meng, "A survey on differential privacy for medical data analysis," *Annals of Data Science*, vol. 11, no. 2, pp. 733–747, 2024.

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

[4] A. Narayanan and V. Shmatikov, "Robust de-anonymization of large sparse datasets," in *2008 IEEE Symposium on Security and Privacy (2008)*, 2008, pp. 111–125.

[5] C. Dwork and A. Roth, "The algorithmic foundations of differential privacy," *Foundations and Trends in Theoretical Computer Science*, vol. 9, no. 3–4, pp. 211–407, 2014. [Online]. Available: http://dx.doi.org/10.1561/0400000042

[6] Apple Research, "Differential privacy overview," 2017, accessed: April 13, 2024. [Online]. Available: https://www.apple.com/privacy/docs/Differential_Privacy_Overview.pdf

[7] Apple Differential Privacy Team, "Learning with privacy at Scale," 2017, accessed: April 13, 2024. [Online]. Available: https://docs-assets.developer.apple.com/ml-research/papers/learning-with-privacy-at-scale.pdf

[8] B. Ding, J. Kulkarni, and S. Yekhanin, "Collecting telemetry data privately," in *Advances in Neural Information Processing Systems*, I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, Eds., vol. 30. Curran Associates, Inc., 2017. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2017/file/253614bbac999b38b5b60cae531c4969-Paper.pdf

[9] B. McMahan and A. Thakurta, "Federated learning with formal differential privacy guarantees," 2022, accessed: April 13, 2024. [Online]. Available: https://blog.research.google/2022/02/federated-learning-with-formal.html

[10] S. Bavadekar, A. Boulanger, J. Davis, D. Desfontaines, E. Gabrilovich, K. Gadepalli, B. Ghazi, T. Griffith, J. Gupta, C. Kamath *et al.*, "Google covid-19 vaccination search insights: Anonymization process description," *arXiv preprint arXiv:2107.01179*, 2021.

[11] A. Aktay, S. Bavadekar, G. Cossoul, J. Davis, D. Desfontaines, A. Fabrikant, E. Gabrilovich, K. Gadepalli, B. Gipson, M. Guevara *et al.*, "Google covid-19 community mobility reports: anonymization process description (version 1.1)," *arXiv preprint arXiv:2004.04145*, 2020.

[12] J. M. Abowd and M. B. Hawes, "Confidentiality protection in the 2020 us census of population and housing," *Annual Review of Statistics and Its Application*, vol. 10, pp. 119–144, 2023.

[13] J. M. Abowd, "The u.s. census bureau adopts differential privacy," in *Proceedings of the 24th ACM SIGKDD international conference on knowledge discovery & data mining*, ser. KDD '18. New York, NY, USA: Association for Computing Machinery, 2018, p. 2867. [Online]. Available: https://doi.org/10.1145/3219819.3226070

[14] S. Ito and M. Terada, "The potential of anonymization method for creating detailed geographical data in japan," *Joint UNECE/Eurostat Work Session on Statistical Data Confidentiality*, vol. 2019, pp. 1–14, 2019.

[15] S. Ito, M. Terada, and S. Kato, "The potential of differential privacy applied to detailed statistical tables created using microdata from the japanese population census," *Expert Meeting on Statistical Data Confidentiality, Conference of European Statisticians (United Nations Economic Commission for Europe)*, September 2023.

[16] R. Rogers, S. Subramaniam, S. Peng, D. Durfee, S. Lee, S. K. Kancha, S. Sahay, and P. Ahammad, "Linkedin's audience engagements api: A privacy preserving data analytics system at scale," *arXiv preprint arXiv:2002.05839*, 2020.

[17] M. Pereira, A. Kim, J. Allen, K. White, J. L. Ferres, and R. Dodhia, "Us broadband coverage data set: a differentially private data release," *arXiv preprint arXiv:2103.14035*, 2021.

[18] M. Paré, M. Teehan, S. Suffian, J. Glass, A. Scheer, and G. M. Young McGee, "Applying energy differential privacy to enable measurement of the ohm connect virtual power plant," 2020, accessed: April 13, 2024. [Online]. Available: https://assets.website-files.com/5cb0a177570549b5f11b9550/5ffddb83b5ea5d67f5c43661_Quantifying%20The%20OhmConnect%20Virtual%20Power%20Plant%20During%20the%20California%20Blackouts.pdf

[19] C. Dwork, N. Kohli, and D. Mulligan, "Differential privacy in practice: Expose your epsilons!" *Journal of Privacy and Confidentiality*, vol. 9, no. 2, 2019. [Online]. Available: https://doi.org/10.29012/jpc.689

[20] M. S. Alvim, M. E. Andrés, K. Chatzikokolakis, P. Degano, and C. Palamidessi, "Differential privacy: on the trade-off between utility and information leakage," in *Formal Aspects of Security and Trust: 8th International Workshop, FAST 2011, Leuven, Belgium, September 12-14, 2011. Revised Selected Papers 8*. Springer, 2012, pp. 39–54.

[21] M. Fredrikson, E. Lantz, S. Jha, S. Lin, D. Page, and T. Ristenpart, "Privacy in pharmacogenetics: An End-to-End case study of personalized warfarin dosing," in *23rd USENIX Security Symposium (USENIX Security 14)*. San Diego, CA: USENIX Association, Aug 2014, pp. 17–32. [Online]. Available: https://www.usenix.org/conference/usenixsecurity14/technical-sessions/presentation/fredrikson_matthew

[22] C. Dwork, "Differential privacy: A survey of results," in *Theory and Applications of Models of Computation*, M. Agrawal, D. Du, Z. Duan, and A. Li, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2008, pp. 1–19.

[23] X. Xiong, S. Liu, D. Li, Z. Cai, and X. Niu, "A comprehensive survey on local differential privacy," *Security and Communication Networks*, vol. 2020, no. 1, pp. 1–29, 2020.

[24] F. K. Dankar and K. El Emam, "Practicing differential privacy in health care: A review." *Trans. Data Priv.*, vol. 6, no. 1, pp. 35–67, 2013.

[25] H. Jiang, J. Pei, D. Yu, J. Yu, B. Gong, and X. Cheng, "Applications of differential privacy in social network analysis: A survey," *IEEE Transactions on Knowledge and Data Engineering*, vol. 35, no. 1, pp. 108–127, 2023.

[26] M. U. Hassan, M. H. Rehmani, and J. Chen, "Differential privacy techniques for cyber physical systems: A survey," *IEEE Communications Surveys & Tutorials*, vol. 22, no. 1, pp. 746–789, 2020.

[27] F. Jin, W. Hua, M. Francia, P. Chao, M. E. Orlowska, and X. Zhou, "A survey and experimental study on privacy-preserving trajectory data publishing," *IEEE Transactions on Knowledge and Data Engineering*, vol. 35, no. 6, pp. 5577–5596, 2023.

[28] N. Ponomareva, S. Vassilvitskii, Z. Xu, B. McMahan, A. Kurakin, and C. Zhang, "How to dp-fy ml: A practical tutorial to machine learning with differential privacy," in *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, ser. KDD '23. New York, NY, USA: Association for Computing Machinery, 2023, p. 5823–5824. [Online]. Available: https://doi.org/10.1145/3580305.3599561

[29] Q. Li, Z. Wen, Z. Wu, S. Hu, N. Wang, Y. Li, X. Liu, and B. He, "A survey on federated learning systems: Vision, hype and reality for data privacy and protection," *IEEE Transactions on Knowledge and Data Engineering*, vol. 35, no. 4, pp. 3347–3366, 2023.

[30] D. Wang, S. Shi, Y. Zhu, and Z. Han, "Federated analytics: Opportunities and challenges," *IEEE Network*, vol. 36, no. 1, pp. 151–158, 2022.

[31] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y. Arcas, "Communication-efficient learning of deep networks from decentralized data," in *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics*, ser. Proceedings of Machine Learning Research, A. Singh and J. Zhu, Eds., vol. 54. PMLR, Apr 2017, pp. 1273–1282. [Online]. Available: https://proceedings.mlr.press/v54/mcmahan17a.html

[32] U. Erlingsson, V. Pihur, and A. Korolova, "Rappor: Randomized aggregatable privacy-preserving ordinal response," in *Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security*, ser. CCS '14. New York, NY, USA: Association for Computing Machinery, 2014, p. 1054–1067.

[33] Z. Jorgensen, T. Yu, and G. Cormode, "Conservative or liberal? personalized differential privacy," in *2015 IEEE 31st International Conference on Data Engineering*, 2015, pp. 1023–1034.

[34] A. Blanco-Justicia, D. Sánchez, J. Domingo-Ferrer, and K. Muralidhar, "A critical review on the use (and misuse) of differential privacy in machine learning," *ACM Comput. Surv.*, vol. 55, no. 8, 2022. [Online]. Available: https://doi.org/10.1145/3547139

[35] B. Balle, J. Bell, A. Gascón, and K. Nissim, "The privacy blanket of the shuffle model," in *Advances in Cryptology – CRYPTO 2019*, A. Boldyreva and D. Micciancio, Eds. Cham: Springer International Publishing, 2019, pp. 638–667.

[36] A. Ghosh, T. Roughgarden, and M. Sundararajan, "Universally utility-maximizing privacy mechanisms," in *Proceedings of the Forty-First Annual ACM Symposium on Theory of Computing*, ser. STOC '09. New York, NY, USA: Association for Computing Machinery, 2009, p. 351–360. [Online]. Available: https://doi.org/10.1145/1536414.1536464

[37] C. L. Canonne, G. Kamath, and T. Steinke, "The discrete gaussian for differential privacy," *Advances in Neural Information Processing Systems*, vol. 33, pp. 15 676–15 688, 2020.

[38] S. L. Warner, "Randomized response: A survey technique for eliminating evasive answer bias," *Journal of the American Statistical Association*, vol. 60, no. 309, pp. 63–69, 1965.

[39] K. Nissim, S. Raskhodnikova, and A. Smith, "Smooth sensitivity and sampling in private data analysis," in *Proceedings of the thirty-ninth annual ACM symposium on Theory of computing*, 2007, pp. 75–84.

[40] J. Zhang, Z. Zhang, X. Xiao, Y. Yang, and M. Winslett, "Functional mechanism: Regression analysis under differential privacy," *arXiv preprint arXiv:1208.0219*, 2012.

[41] C. Dwork and J. Lei, "Differential privacy and robust statistics," in *Proceedings of the Forty-First Annual ACM Symposium on Theory of Computing*, ser. STOC '09. New York, NY, USA: Association for Computing Machinery, 2009, p. 371–380. [Online]. Available: https://doi.org/10.1145/1536414.1536466

[42] I. Mironov, "Rényi differential privacy," in *2017 IEEE 30th Computer Security Foundations Symposium (CSF)*, 2017, pp. 263–275.

[43] C. Dwork and G. N. Rothblum, "Concentrated differential privacy," *arXiv preprint arXiv:1603.01887*, 2016.

[44] M. Bun and T. Steinke, "Concentrated differential privacy: Simplifications, extensions, and lower bounds," in *Theory of Cryptography*, M. Hirt and A. Smith, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2016, pp. 635–658.

[45] M. Bun, C. Dwork, G. N. Rothblum, and T. Steinke, "Composable and versatile privacy via truncated cdp," in *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing*, ser. STOC 2018. New York, NY, USA: Association for Computing Machinery, 2018, p. 74–86. [Online]. Available: https://doi.org/10.1145/3188745.3188946

[46] A. Rényi, "On measures of entropy and information," in *Proceedings of the fourth Berkeley symposium on mathematical statistics and probability, volume 1: contributions to the theory of statistics*, vol. 4. University of California Press, 1961, pp. 547–562.

[47] I. Mironov, K. Talwar, and L. Zhang, "Rényi differential privacy of the sampled gaussian mechanism," *arXiv preprint arXiv:1908.10530*, 2019.

[48] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas, "Communication-efficient learning of deep networks from decentralized data," in *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics*, vol. 54. PMLR, 2017, pp. 1273–1282.

[49] K. Chaudhuri and C. Monteleoni, "Privacy-preserving logistic regression," *Advances in neural information processing systems*, vol. 21, 2008.

[50] J. Vaidya, H. Yu, and X. Jiang, "Privacy-preserving svm classification," *Knowledge and Information Systems*, vol. 14, pp. 161–178, 2008.

[51] M. Abadi, A. Chu, I. Goodfellow, H. B. McMahan, I. Mironov, K. Talwar, and L. Zhang, "Deep learning with differential privacy," in *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*, ser. CCS '16. New York, NY, USA: Association for Computing Machinery, 2016, p. 308–318. [Online]. Available: https://doi.org/10.1145/2976749.2978318

[52] H. B. McMahan, D. Ramage, K. Talwar, and L. Zhang, "Learning differentially private recurrent language models," *International Conference on Learning Representations (ICLR)*, 2018.

[53] D. Desfontaines, "A list of real-world uses of differential privacy," https://desfontain.es/privacy/real-world-differential-privacy.html, 10 2021, ted is writing things (personal blog), accessed: April 13, 2024.

[54] R. Helmer, A. Miyaguchi, and E. Rescorla. (2018, Oct.) Testing privacy-preserving telemetry with Prio. Mozilla Hacks. [Online]. Available: https://hacks.mozilla.org/2018/10/testing-privacy-preserving-telemetry-with-prio/

[55] P. Kairouz, B. Mcmahan, S. Song, O. Thakkar, A. Thakurta, and Z. Xu, "Practical and private (deep) learning without sampling or shuffling," in *Proceedings of the 38th International Conference on Machine Learning*, ser. Proceedings of Machine Learning Research, M. Meila and T. Zhang, Eds., vol. 139. PMLR, 18–24 Jul 2021, pp. 5213–5225. [Online]. Available: https://proceedings.mlr.press/v139/kairouz21b.html

[56] Z. Xu, Y. Zhang, G. Andrew, C. A. Choquette-Choo, P. Kairouz, H. B. McMahan, J. Rosenstock, and Y. Zhang, "Federated learning of gboard language models with differential privacy," *arXiv preprint arXiv:2305.18465*, 2023.

[57] B. Zhang, V. Doroshenko, P. Kairouz, T. Steinke, A. Thakurta, Z. Ma, E. Cohen, H. Apte, and J. Spacek, "Differentially private stream processing at scale," *arXiv preprint arXiv:2303.18086*, 2023.

[58] H. Corrigan-Gibbs and D. Boneh, "Prio: Private, robust, and scalable computation of aggregate statistics," in *14th USENIX Symposium on Networked Systems Design and Implementation (NSDI 17)*. Boston, MA: USENIX Association, March 2017, pp. 259–282. [Online]. Available: https://www.usenix.org/conference/nsdi17/technical-sessions/presentation/corrigan-gibbs

[59] Apple Machine Learning Research, "Scenes: Differential Privacy in Action," 2023, accessed: February 05, 2025. [Online]. Available: https://machinelearning.apple.com/research/scenes-differential-privacy

[60] T. Adeleye, S. Berghel, D. Desfontaines, M. Hay, I. Johnson, C. Lemoisson, A. Machanavajjhala, T. Magerlein, G. Modena, D. Pujol *et al.*, "Publishing wikipedia usage data with strong privacy guarantees," *arXiv preprint arXiv:2308.16298*, 2023.

[61] Wikimedia Foundation, "Geoeditors Monthly Dataset," 2024, accessed: February 05, 2025. [Online]. Available: https://analytics.wikimedia.org/published/datasets/geoeditors_monthly/00_README.html

[62] ——, "Russian Editor Information (2022-23)," 2024, accessed: February 05, 2025. [Online]. Available: https://wikitech.wikimedia.org/wiki/Russian_editor_information_(2022-23)

[63] Rogers, Ryan and Subramaniam, Subbu and Xu, Lin, "Privacy-Preserving Single Post Analytics," 2023, accessed: February 05, 2025. [Online]. Available: https://www.linkedin.com/blog/engineering/trust-and-safety/privacy-preserving-single-post-analytics

[64] S. Messing, C. DeGregorio, B. Hillenbrand, G. King, S. Mahanti, Z. Mukerjee, C. Nayak, N. Persily, A. Wilkins *et al.*, "Facebook privacy-protected full urls data set," *Harvard Dataverse*, 2020. [Online]. Available: https://dataverse.harvard.edu/file.xhtml?persistentId=doi:10.7910/DVN/TDOAPG/DGSAMS

[65] R. Microsoft, "Assistive ai makes replying easier," 2021, accessed: April 13, 2024. [Online]. Available: https://www.microsoft.com/en-us/research/articles/assistive-ai-makes-replying-easier-2/

[66] J. Konečnỳ, H. B. McMahan, D. Ramage, and P. Richtárik, "Federated optimization: Distributed machine learning for on-device intelligence," *arXiv preprint arXiv:1610.02527*, 2016.

[67] F. Hartmann and P. Kairouz, "Predicting text selections with federated learning," 2021, accessed: April 13, 2024. [Online]. Available: https://research.google/blog/predicting-text-selections-with-federated-learning/

[68] A. Bassolas, H. Barbosa-Filho, B. Dickinson, X. Dotiwalla, P. Eastham, R. Gallotti, G. Ghoshal, B. Gipson, S. A. Hazarie, H. Kautz *et al.*, "Hierarchical organization of urban mobility and its connection with city livability," *Nature communications*, vol. 10, no. 1, p. 4817, 2019.

[69] A. Sadilek and X. Dotiwalla, "New insights into human mobility with privacy preserving aggregation," 2019, accessed: April 13, 2024. [Online]. Available: https://blog.research.google/2019/11/new-insights-into-human-mobility-with.html

[70] C. Bian, A. Cheu, Y. Guzman, M. Gruteser, P. Kairouz, R. McKenna, and E. Roth, "Releasing large-scale human mobility histograms with differential privacy," *arXiv preprint arXiv:2407.03496*, 2024.

[71] K. Tezapsidis, "Uber releases open source project for differential privacy," https://medium.com/uber-security-privacy/differential-privacy-open-source-7892c82c42b6, Jul. 2017, uber Privacy and Security (blog).

[72] N. Johnson, J. P. Near, J. M. Hellerstein, and D. Song, "Chorus: a programming framework for building scalable differential privacy mechanisms," in *2020 IEEE European Symposium on Security and Privacy (EuroS&P)*, 2020, pp. 535–551.

[73] U.S. Department of Health and Human Services, "Health insurance portability and accountability act of 1996," 1996, accessed: April 13, 2024. [Online]. Available: https://aspe.hhs.gov/reports/health-insurance-portability-accountability-act-1996

[74] Spectus, "Differential privacy implementation for hurricane irma dashboard," 2022, accessed: April 13, 2024.

[75] S. Bavadekar, A. Dai, J. Davis, D. Desfontaines, I. Eckstein, K. Everett, A. Fabrikant, G. Flores, E. Gabrilovich, K. Gadepalli *et al.*, "Google covid-19 search trends symptoms dataset: Anonymization process description (version 1.0)," *arXiv preprint arXiv:2009.01265*, 2020.

[76] Apple and Google, "Exposure notification privacy-preserving analytics (enpa), white paper," April 2021, accessed: April 13, 2024. [Online]. Available: https://covid19-static.cdn-apple.com/applications/covid19/current/static/contact-tracing/pdf/ENPA_White_Paper.pdf

[77] A. Machanavajjhala, D. Kifer, J. Abowd, J. Gehrke, and L. Vilhuber, "Privacy: Theory meets practice on the map," in *2008 IEEE 24th International Conference on Data Engineering*, 2008, pp. 277–286.

[78] A. D. Foote, A. Machanavajjhala, and K. McKinney, "Releasing earnings distributions using differential privacy: Disclosure avoidance system for post-secondary employment outcomes (pseo)," *Journal*

of Privacy and Confidentiality, vol. 9, no. 2, Oct. 2019. [Online]. Available: https://journalprivacyconfidentiality.org/index.php/jpc/article/view/722

[79] C. T. Kenny, S. Kuriwaki, C. McCartan, E. T. Rosenman, T. Simko, and K. Imai, "The use of differential privacy for census data and its impact on redistricting: The case of the 2020 us census," *Science advances*, vol. 7, no. 41, p. eabk3283, 2021.

[80] A. Sadilek and X. Dotiwalla, "Researching formal privacy for the census bureau's county business patterns program," 2023, accessed: April 13, 2024. [Online]. Available: https://www.census.gov/topics/business-economy/disclosure/about.html

[81] G. Miklau, "How tumult labs helped the irs support educational accountability with differential privacy," July 2022, accessed: April 13, 2024. [Online]. Available: https://www.tmlt.io/casestudy/illuminating-college-outcomes-while-protecting-privacy

[82] S. Berghel, P. Bohannon, D. Desfontaines, C. Estes, S. Haney, L. Hartman, M. Hay, A. Machanavajjhala, T. Magerlein, G. Miklau *et al.*, "Tumult analytics: a robust, easy-to-use, scalable, and expressive framework for differential privacy," *arXiv preprint arXiv:2212.04133*, 2022.

[83] I. Dove, "Applying differential privacy protection to ons mortality data, pilot study," August 2021, accessed: April 13, 2024. [Online]. Available: https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/deaths/methodologies/applyingdifferentialprivacyprotectiontoonsmortalitydatapilotstudy

[84] J. Bailie and C.-H. Chien, "Abs perturbation methodology through the lens of differential privacy," *UN Economic Commission for Europe, Work Session on Statistical Data Confidentiality*, 2019.

[85] S. Hod and R. Canetti, "Differentially Private Release of Israel's National Registry of Live Births," in *2025 IEEE Symposium on Security and Privacy (SP)*. Los Alamitos, CA, USA: IEEE Computer Society, May 2025, pp. 3912–3930. [Online]. Available: https://doi.ieeecomputersociety.org/10.1109/SP61157.2025.00101

[86] R. Rogers, A. R. Cardoso, K. Mancuhan, A. Kaura, N. Gahlawat, N. Jain, P. Ko, and P. Ahammad, "A members first approach to enabling linkedin's labor market insights at scale," *arXiv preprint arXiv:2010.13981*, 2020.

[87] S. Badrinarayanan, O. Osoba, M. Cheng, R. Rogers, S. Jain, R. Tandra, and N. S. Pillai, "Privacy-preserving race/ethnicity estimation for algorithmic bias measurement in the us," *arXiv preprint arXiv:2409.04652*, 2024.

[88] Y. Xiao, L. Xiong, L. Fan, and S. Goryczka, "Dpcube: Differentially private histogram release through multidimensional partitioning," *arXiv preprint arXiv:1202.5358*, 2012.

[89] G. Ács and C. Castelluccia, "I have a dream! (differentially private smart metering)," in *Information Hiding*, T. Filler, T. Pevný, S. Craver, and A. Ker, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, pp. 118–132.

[90] L. Ou, Z. Qin, S. Liao, T. Li, and D. Zhang, "Singular spectrum analysis for local differential privacy of classifications in the smart grid," *IEEE Internet of Things Journal*, vol. 7, no. 6, pp. 5246–5255, 2020.

[91] Q. Wang, Y. Zhang, X. Lu, Z. Wang, Z. Qin, and K. Ren, "Rescuedp: Real-time spatio-temporal crowd-sourced data publishing with differential privacy," in *IEEE INFOCOM 2016 - The 35th Annual IEEE International Conference on Computer Communications*, 2016, pp. 1–9.

[92] G. Eibl and D. Engel, "Differential privacy for real smart metering data," *Computer Science-Research and Development*, vol. 32, pp. 173–182, 2017.

[93] R. Chen, H. Li, A. K. Qin, S. P. Kasiviswanathan, and H. Jin, "Private spatial data aggregation in the local setting," in *2016 IEEE 32nd International Conference on Data Engineering (ICDE)*, 2016, pp. 289–300.

[94] F. McSherry and R. Mahajan, "Differentially-private network trace analysis," in *Proceedings of the ACM SIGCOMM 2010 Conference*, ser. SIGCOMM '10. New York, NY, USA: Association for Computing Machinery, 2010, pp. 123–134. [Online]. Available: https://doi.org/10.1145/1851182.1851199

[95] V. Ruehle, R. Sim, S. Yekhanin, N. Chandran, M. Chase, Jones *et al.*, "Privacy preserving machine learning: Maintaining confidentiality and preserving trust," 2021, accessed: May 13, 2025. [Online]. Available: https://www.microsoft.com/en-us/research/blog/privacy-preserving-machine-learning-maintaining-confidentiality-and-preserving-trust/

[96] P. Barbosa, A. Brito, and H. Almeida, "A technique to provide differential privacy for appliance usage in smart metering," *Information Sciences*, vol. 370-371, pp. 355–367, 2016. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0020025516305862

[97] G. M. Garrido, X. Liu, F. Matthes, and D. Song, "Lessons learned: Surveying the practicality of differential privacy in the industry," *arXiv preprint arXiv:2211.03898*, 2022.

[98] P. G. Kelley, J. Bresee, L. F. Cranor, and R. W. Reeder, "A "nutrition label" for privacy," in *Proceedings of the 5th Symposium on Usable Privacy and Security (SOUPS '09)*, ser. SOUPS '09. New York, NY, USA: Association for Computing Machinery, 2009. [Online]. Available: https://doi.org/10.1145/1572532.1572538

[99] S. Benthall and R. Cummings, "Integrating differential privacy and contextual integrity," in *Proceedings of the Symposium on Computer Science and Law*, ser. CSLAW '24. New York, NY, USA: Association for Computing Machinery, 2024, p. 9–15.

[100] J. Sarathy, S. Song, A. Haque, T. Schlatter, and S. Vadhan, "Don't look at the data! how differential privacy reconfigures the practices of data science," in *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, 2023, pp. 1–19.

[101] I. Rosenberg, A. Shabtai, Y. Elovici, and L. Rokach, "Adversarial machine learning attacks and defense methods in the cyber security domain," *ACM Comput. Surv.*, vol. 54, no. 5, may 2021. [Online]. Available: https://doi.org/10.1145/3453158

[102] R. Bitton, N. Maman, I. Singh, S. Momiyama, Y. Elovici, and A. Shabtai, "Evaluating the cybersecurity risk of real-world, machine learning production systems," *ACM Comput. Surv.*, vol. 55, no. 9, jan 2023. [Online]. Available: https://doi.org/10.1145/3559104

[103] Z. Sun and H. Sun, "Improving gboard language models via private federated analytics," April 2024, accessed: April 13, 2024. [Online]. Available: https://research.google/blog/improving-gboard-language-models-via-private-federated-analytics/

[104] H. Eichner, D. Ramage, K. Bonawitz, D. Huba, T. Santoro, B. McLarnon, T. Van Overveldt, N. Fallen, P. Kairouz, A. Cheu *et al.*, "Confidential federated computations," *arXiv preprint arXiv:2404.10764*, 2024.

[105] Z. Sun, P. Kairouz, H. Sun, A. Gascon, and A. T. Suresh, "Private federated discovery of out-of-vocabulary words for gboard," *arXiv preprint arXiv:2404.11607*, 2024.

[106] Y. Zhang, D. Ramage, Z. Xu, Y. Zhang, S. Zhai, and P. Kairouz, "Private federated learning in gboard," *arXiv preprint arXiv:2306.14793*, 2023.

[107] Z. Xu and Y. Zhang, "Advances in private training for production on-device language models," February 2024, accessed: April 13, 2024. [Online]. Available: https://research.google/blog/advances-in-private-training-for-production-on-device-language-models/

[108] A. Herdağdelen, A. Dow, S. Bogdan, M. Payman, and A. Pompe, "Protecting privacy in facebook mobility data during the covid-19 response," *Facebook Research*, 2020. [Online]. Available: https://research.facebook.com/blog/2020/06/protecting-privacy-in-facebook-mobility-data-during-the-covid-19-response/

[109] S. Gopi, S. Mahabadi, and S. Yekhanin, "The global victim-perpetrator synthetic dataset," 2023, accessed: April 13, 2024. [Online]. Available: https://www.ctdatacollaborative.org/global-victim-perpetrator-synthetic-dataset

[110] Microsoft Research. (2022, Dec.) Iom and microsoft release first-ever differentially private synthetic dataset to counter human trafficking. Microsoft Research Blog. [Online]. Available: https://www.microsoft.com/en-us/research/blog/iom-and-microsoft-release-first-ever-differentially-private-synthetic-dataset-to-counter-human-trafficking/

[111] A. Foote, H. Joyce, S. Tibbets, and L. Warren, "Post-secondary employment outcomes (pseo)," 2023, accessed: April 13, 2024. [Online]. Available: https://lehd.ces.census.gov/doc/PSEOTechnicalDocumentation.pdf

[112] F. Andersson, J. M. Abowd, M. Graham, J. Wu, and L. Vilhuber, "Formal privacy guarantees and analytical validity of onthemap public-use data," *NSF-Census-IRS Workshop on Synthetic Data and Confidentiality Protection*, 2009. [Online]. Available: https://ecommons.cornell.edu/server/api/core/bitstreams/253b25f5-5cc3-400b-825a-a0504b2db58c/content

[113] Spectus, "Hurricane irma dashboard," 2022, accessed: April 13, 2024. [Online]. Available: https://cuebiq.com/social-impact/

[114] N. Johnson, J. P. Near, and D. Song, "Towards practical differential privacy for sql queries," *Proceedings of the VLDB Endowment*, vol. 11, no. 5, pp. 526–539, 2018.

[115] A. Korolova, K. Kenthapadi, N. Mishra, and A. Ntoulas, "Releasing search queries and clicks privately," in *Proceedings of the 18th International Conference on World Wide Web*, ser. WWW '09. New York, NY, USA: Association for Computing Machinery, 2009, p. 171–180. [Online]. Available: https://doi.org/10.1145/1526709.1526733

[116] D. Vu and A. Slavkovic, "Differential privacy for clinical trial data: Preliminary evaluations," in *2009 IEEE International Conference on Data Mining Workshops*. IEEE, 2009, pp. 138–143.

[117] P. Kodeswaran and E. Viegas, "Applying differential privacy to search queries in a policy based interactive framework," in *Proceedings of the ACM first international workshop on Privacy and anonymity for very large databases*, 2009, pp. 25–32.

[118] R. Bhaskar, S. Laxman, A. Smith, and A. Thakurta, "Discovering frequent patterns in sensitive data," in *Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ser. KDD '10. New York, NY, USA: Association for Computing Machinery, 2010, p. 503–512. [Online]. Available: https://doi.org/10.1145/1835804.1835869

[119] M. Pathak, S. Rane, and B. Raj, "Multiparty differential privacy via aggregation of locally trained classifiers," *Advances in neural information processing systems*, vol. 23, 2010.

[120] A. Machanavajjhala, A. Korolova, and A. D. Sarma, "Personalized social recommendations-accurate or private?" *arXiv preprint arXiv:1105.4254*, 2011.

[121] R. Chen, B. Fung, and B. C. Desai, "Differentially private trajectory data publication," *arXiv preprint arXiv:1112.2020*, 2011.

[122] G. Cormode, C. Procopiuc, D. Srivastava, E. Shen, and T. Yu, "Differentially private spatial decompositions," in *2012 IEEE 28th International Conference on Data Engineering*, 2012, pp. 20–31.

[123] L. Bonomi, L. Xiong, R. Chen, and B. Fung, "Privacy preserving record linkage via grams projections," *arXiv preprint arXiv:1208.2773*, 2012.

[124] A. Narayan and A. Haeberlen, "DJoin: Differentially private join queries over distributed databases," in *10th USENIX Symposium on Operating Systems Design and Implementation (OSDI 12)*. Hollywood, CA: USENIX Association, 2012, pp. 149–162. [Online]. Available: https://www.usenix.org/conference/osdi12/technical-sessions/presentation/narayan

[125] R. Chen, A. Reznichenko, P. Francis, and J. Gehrke, "Towards statistical queries over distributed private user data," in *9th USENIX Symposium on Networked Systems Design and Implementation (NSDI 12)*. San Jose, CA: USENIX Association, 2012, pp. 169–182. [Online]. Available: https://www.usenix.org/conference/nsdi12/technical-sessions/presentation/chen_ruichuan

[126] C. Li and G. Miklau, "An adaptive mechanism for accurate query answering under differential privacy," *arXiv preprint arXiv:1202.3807*, 2012.

[127] S. A. Vinterbo, A. D. Sarwate, and A. A. Boxwala, "Protecting count queries in study design," *Journal of the American Medical Informatics Association*, vol. 19, no. 5, pp. 750–757, 04 2012. [Online]. Available: https://doi.org/10.1136/amiajnl-2011-000459

[128] N. Mohammed, X. Jiang, R. Chen, B. C. M. Fung, and L. Ohno-Machado, "Privacy-preserving heterogeneous health data sharing," *Journal of the American Medical Informatics Association*, vol. 20, no. 3, pp. 462–469, 12 2012. [Online]. Available: https://doi.org/10.1136/amiajnl-2012-001027

[129] K. Chaudhuri, A. D. Sarwate, and K. Sinha, "A near-optimal algorithm for differentially-private principal components." *Journal of Machine Learning Research*, vol. 14, 2013.

[130] C. Uhlerop, A. Slavković, and S. E. Fienberg, "Privacy-preserving data sharing for genome-wide association studies," *The Journal of privacy and confidentiality*, vol. 5, no. 1, p. 137, 2013.

[131] 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*, ser. CCS '13. New York, NY, USA: Association for Computing Machinery, 2013, p. 901–914. [Online]. Available: https://doi.org/10.1145/2508859.2516735

[132] Y. Wang, X. Wu, and L. Wu, "Differential privacy preserving spectral graph analysis," in *Advances in Knowledge Discovery and Data Mining: 17th Pacific-Asia Conference, PAKDD 2013, Gold Coast, Australia, April 14-17, 2013, Proceedings, Part II 17*. Springer, 2013, pp. 329–340.

[133] M. Backes and S. Meiser, "Differentially private smart metering with battery recharging," in *International Workshop on Data Privacy Management*. Springer, 2013, pp. 194–212.

[134] L. Fan, L. Bonomi, L. Xiong, and V. Sunderam, "Monitoring web browsing behavior with differential privacy," in *Proceedings of the 23rd International Conference on World Wide Web*, ser. WWW '14. New York, NY, USA: Association for Computing Machinery, 2014, p. 177–188. [Online]. Available: https://doi.org/10.1145/2566486.2568038

[135] W. Lu and G. Miklau, "Exponential random graph estimation under differential privacy," in *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ser. KDD '14. New York, NY, USA: Association for Computing Machinery, 2014, p. 921–930. [Online]. Available: https://doi.org/10.1145/2623330.2623683

[136] Y. Shen and H. Jin, "Privacy-preserving personalized recommendation: An instance-based approach via differential privacy," in *2014 IEEE international conference on data mining*. IEEE, 2014, pp. 540–549.

[137] Y. Xiao and L. Xiong, "Protecting locations with differential privacy under temporal correlations," in *Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security*, ser. CCS '15. New York, NY, USA: Association for Computing Machinery, 2015, p. 1298–1309. [Online]. Available: https://doi.org/10.1145/2810103.2813640

[138] H. Li, Y. Dai, and X. Lin, "Efficient e-health data release with consistency guarantee under differential privacy," in *2015 17th International Conference on E-health Networking, Application & Services (HealthCom)*, 2015, pp. 602–608.

[139] S. Han, S. Zhao, Q. Li, C.-H. Ju, and W. Zhou, "Ppm-hda: Privacy-preserving and multifunctional health data aggregation with fault tolerance," *IEEE Transactions on Information Forensics and Security*, vol. 11, no. 9, pp. 1940–1955, 2016.

[140] N. Phan, Y. Wang, X. Wu, and D. Dou, "Differential privacy preservation for deep auto-encoders: an application of human behavior prediction," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 30, no. 1, Feb. 2016.

[141] C. Lin, Z. Song, H. Song, Y. Zhou, Y. Wang, and G. Wu, "Differential privacy preserving in big data analytics for connected health," *Journal of medical systems*, vol. 40, pp. 1–9, 2016.

[142] Y. Shen and H. Jin, "Epicrec: Towards practical differentially private framework for personalized recommendation," in *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*, ser. CCS '16. New York, NY, USA: Association for Computing Machinery, 2016, p. 180–191. [Online]. Available: https://doi.org/10.1145/2976749.2978316

[143] T. T. Nguyên, X. Xiao, Y. Yang, S. C. Hui, H. Shin, and J. Shin, "Collecting and analyzing data from smart device users with local differential privacy," *arXiv preprint arXiv:1606.05053*, 2016.

[144] Z. Qin, T. Yu, Y. Yang, I. Khalil, X. Xiao, and K. Ren, "Generating synthetic decentralized social graphs with local differential privacy," in *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*, ser. CCS '17. New York, NY, USA: Association for Computing Machinery, 2017, p. 425–438. [Online]. Available: https://doi.org/10.1145/3133956.3134086

[145] C. Yin, J. Xi, R. Sun, and J. Wang, "Location privacy protection based on differential privacy strategy for big data in industrial internet of things," *IEEE Transactions on Industrial Informatics*, vol. 14, no. 8, pp. 3628–3636, 2017.

[146] J. Chen, H. Ma, D. Zhao, and L. Liu, "Correlated differential privacy protection for mobile crowdsensing," *IEEE Transactions on Big Data*, vol. 7, no. 4, pp. 784–795, 2017.

[147] A. Bittau, U. Erlingsson, P. Maniatis, I. Mironov, A. Raghunathan, D. Lie, M. Rudominer, U. Kode, J. Tinnes, and B. Seefeld, "Prochlo: Strong privacy for analytics in the crowd," in *Proceedings of the 26th Symposium on Operating Systems Principles*, ser. SOSP '17. New York, NY, USA: Association for Computing Machinery, 2017, p. 441–459. [Online]. Available: https://doi.org/10.1145/3132747.3132769

[148] X. Xiong, F. Chen, P. Huang, M. Tian, X. Hu, B. Chen, and J. Qin, "Frequent itemsets mining with differential privacy over large-scale data," *IEEE Access*, vol. 6, pp. 28 877–28 889, 2018.

[149] T. Zhu, M. Yang, P. Xiong, Y. Xiang, and W. Zhou, "An iteration-based differentially private social network data release," *Computer Systems Science and Engineering*, vol. 33, no. 2, pp. 61–69, 2018.

[150] L. Lyu, K. Nandakumar, B. Rubinstein, J. Jin, J. Bedo, and M. Palaniswami, "Ppfa: Privacy preserving fog-enabled aggregation in smart grid," *IEEE Transactions on Industrial Informatics*, vol. 14, no. 8, pp. 3733–3744, 2018.

[151] M. Khavkin and M. Last, "Preserving differential privacy and utility of non-stationary data streams," in *2018 IEEE International Conference on Data Mining Workshops (ICDMW)*, 2018, pp. 29–34.

[152] L. Fan, "Image pixelization with differential privacy," in *Data and Applications Security and Privacy XXXII*, Springer. Springer International Publishing, 2018, pp. 148–162.

[153] J. W. Kim, D.-H. Kim, and B. Jang, "Application of local differential privacy to collection of indoor positioning data," *IEEE Access*, vol. 6, pp. 4276–4286, 2018.

[154] W.-S. Choi, M. Tomei, J. R. S. Vicarte, P. K. Hanumolu, and R. Kumar, "Guaranteeing local differential privacy on ultra-low-power systems," in *2018 ACM/IEEE 45th Annual International Symposium on Computer Architecture (ISCA)*, 2018, pp. 561–574.

[155] M. Asada, M. Yoshikawa, and Y. Cao, ""when and where do you want to hide?" – recommendation of location privacy preferences with local differential privacy," in *Data and Applications Security and Privacy XXXIII*, S. N. Foley, Ed. Cham: Springer International Publishing, 2019, pp. 164–176.

[156] R. J. Wilson, C. Y. Zhang, W. Lam, D. Desfontaines, D. Simmons-Marengo, and B. Gipson, "Differentially private sql with bounded user contribution," *arXiv preprint arXiv:1909.01917*, 2019.

[157] X. Zhao, Y. Li, Y. Yuan, X. Bi, and G. Wang, "Ldpart: Effective location-record data publication via local differential privacy," *IEEE Access*, vol. 7, pp. 31 435–31 445, 2019.

[158] N. Fernandes, M. Dras, and A. McIver, "Generalised differential privacy for text document processing," in *Principles of Security and Trust: 8th International Conference, POST 2019, Held as Part of the European Joint Conferences on Theory and Practice of Software, ETAPS 2019, Prague, Czech Republic, 2019, Proceedings 8*. Springer International Publishing, 2019, pp. 123–148.

[159] H. Wang, S. Xie, and Y. Hong, "Videodp: A universal platform for video analytics with differential privacy," *arXiv preprint arXiv:1909.08729*, 2019.

[160] J. Steil, I. Hagestedt, M. X. Huang, and A. Bulling, "Privacy-aware eye tracking using differential privacy," in *Proceedings of the 11th ACM Symposium on Eye Tracking Research & Applications*, ser. ETRA '19. New York, NY, USA: Association for Computing Machinery, 2019. [Online]. Available: https://doi.org/10.1145/3314111.3319915

[161] H. Cho, S. Simmons, R. Kim, and B. Berger, "Privacy-preserving biomedical database queries with optimal privacy-utility trade-offs," *Cell systems*, vol. 10, no. 5, pp. 408–416, 2020.

[162] H. Lee and Y. D. Chung, "Differentially private release of medical microdata: an efficient and practical approach for preserving informative attribute values," *BMC Medical Informatics and Decision Making*, vol. 20, no. 1, pp. 1–15, 2020.

[163] M. Chamikara, P. Bertok, I. Khalil, D. Liu, and S. Camtepe, "Privacy preserving face recognition utilizing differential privacy," *Computers & Security*, vol. 97, p. 101951, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0167404820302273

[164] Z. Wang, H. Guo, Z. Zhang, M. Song, S. Zheng, Q. Wang, and B. Niu, "Towards compression-resistant privacy-preserving photo sharing on social networks," in *Proceedings of the Twenty-First International Symposium on Theory, Algorithmic Foundations, and Protocol Design for Mobile Networks and Mobile Computing*, ser. Mobihoc '20. New York, NY, USA: Association for Computing Machinery, 2020, p. 81–90. [Online]. Available: https://doi.org/10.1145/3397166.3409141

[165] M. Sun, Q. Wang, and Z. Liu, "Human action image generation with differential privacy," in *2020 IEEE International Conference on Multimedia and Expo (ICME)*, 2020, pp. 1–6.

[166] J. Lin, J. Niu, X. Liu, and M. Guizani, "Protecting your shopping preference with differential privacy," *IEEE Transactions on Mobile Computing*, vol. 20, no. 5, pp. 1965–1978, 2021.

[167] J. Qian, H. Du, J. Hou, L. Chen, T. Jung, and X.-Y. Li, "Speech sanitizer: Speech content desensitization and voice anonymization," *IEEE Transactions on Dependable and Secure Computing*, vol. 18, no. 6, pp. 2631–2642, 2021.

[168] 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*. Association for Computational Linguistics, aug 2021, pp. 3853–3866. [Online]. Available: https://aclanthology.org/2021.findings-acl.337.pdf

[169] T. Cunningham, G. Cormode, H. Ferhatosmanoglu, and D. Srivastava, "Real-world trajectory sharing with local differential privacy," *arXiv preprint arXiv:2108.02084*, 2021.

[170] J. Han and S. Cai, "Differentially private task allocation algorithm under preference protection," *IEEE Access*, vol. 10, pp. 33 059–33 068, 2022.

[171] X. Chen, S. Miao, and Y. Wang, "Differential privacy in personalized pricing with nonparametric demand models," *Operations Research*, vol. 71, no. 2, pp. 581–602, 2023.

[172] B. Chen, K. Leahy, A. Jones, and M. Hale, "Differential privacy for symbolic systems with application to markov chains," *Automatica*, vol. 152, p. 110908, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0005109823000584

[173] X. Sun, Q. Ye, H. Hu, Y. Wang, K. Huang, T. Wo, and J. Xu, "Synthesizing realistic trajectory data with differential privacy," *IEEE Transactions on Intelligent Transportation Systems*, vol. 24, no. 5, pp. 5502–5515, 2023.

[174] Z. Xu, M. Collins, Y. Wang, L. Panait, S. Oh, S. Augenstein, T. Liu, F. Schroff, and H. B. McMahan, "Learning to generate image embeddings with user-level differential privacy," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2023, pp. 7969–7980.

[175] J. Chen, J. Xue, Y. Wang, L. Huang, T. Baker, and Z. Zhou, "Privacy-preserving and traceable federated learning for data sharing in industrial iot applications," *Expert Systems with Applications*, vol. 213, p. 119036, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0957417422020541

[176] J. Acharya, Y. Liu, and Z. Sun, "Discrete distribution estimation under user-level local differential privacy," in *Proceedings of The 26th International Conference on Artificial Intelligence and Statistics*, ser. Proceedings of Machine Learning Research, F. Ruiz, J. Dy, and J.-W. van de Meent, Eds., vol. 206. PMLR, 25–27 Apr 2023, pp. 8561–8585. [Online]. Available: https://proceedings.mlr.press/v206/acharya23a.html

[177] H. Hu, G. Dobbie, Z. Salcic, M. Liu, J. Zhang, L. Lyu, and X. Zhang, "Differentially private locality sensitive hashing based federated recommender system," *Concurrency and Computation: Practice and Experience*, vol. 35, no. 14, pp. 4075–4096, 2023.

[178] D. Keeler, C. Komlo, E. Lepert, S. Veitch, and X. He, "Dprio: Efficient differential privacy with high utility for prio," *Proceedings on Privacy Enhancing Technologies*, no. 3, pp. 375–390, 2023.

[179] H. Batool, A. Anjum, A. Khan, S. Izzo, C. Mazzocca, and G. Jeon, "A secure and privacy preserved infrastructure for vanets based on federated learning with local differential privacy," *Information Sciences*, vol. 652, p. 119717, 2024.

[180] A. Qashlan, P. Nanda, and M. Mohanty, "Differential privacy model for blockchain based smart home architecture," *Future Generation Computer Systems*, vol. 150, pp. 49–63, 2024.

[181] D. Novado, E. Cohen, and J. Foster, "Multi-tier privacy protection for large language models using differential privacy," *Authorea Preprints*, 2024.

[182] Z. Ju and Y. Li, "Vehicle-to-vehicle energy sharing scheme: A privacy-preserving solution based on local differential privacy method," *IEEE Network*, 2024.

[183] Y. Shanmugarasa, M. Chamikara, H.-y. Paik, S. S. Kanhere, and L. Zhu, "Local differential privacy for smart meter data sharing with energy disaggregation," in *2024 20th International Conference on Distributed Computing in Smart Systems and the Internet of Things (DCOSS-IoT)*. IEEE, 2024, pp. 1–10.

[184] V. Agrawal, S. V. Kalmady, V. M. Manoj, M. V. Manthena, W. Sun, M. S. Islam, A. Hindle, P. Kaul, and R. Greiner, "Federated learning and differential privacy techniques on multi-hospital population-scale electrocardiogram data," in *Proceedings of the 2024 8th International Conference on Medical and Health Informatics*, 2024, pp. 143–152.

Michael Khavkin is a PhD candidate in the School of Industrial & Intelligent Systems Engineering at Tel Aviv University. His research interests include machine learning, usable security and the intersection of Human-Computer Interaction (HCI) and privacy-preserving data analysis with Differential Privacy.

Eran Toch is an Associate Professor at the Faculty of Engineering at Tel-Aviv University, where he heads the School of Industrial & Intelligent Systems Engineering. He co-directs the Interacting with Technology Lab (IWiT), working on areas such as usable security and privacy and human-AI interaction.

APPENDIX

TABLE V
SURVEYED DP APPLICATIONS IN ACADEMIC, COMMERCIAL, AND GOVERNMENTAL/NON-PROFIT SETTINGS.

Reference	Year	Setting	Data size ($\times 10^3$)	Employed ε
Chaudhuri et al. [49]	2008	Acad.	17.5	$[0.01, 0.2]$
Machanavajjhala et al. [77]	2008	Govt.	1500	8.6
Korolova et al. [115]	2009	Acad.	$490 \cdot 10^3$	$[0.7, 2.3]$
Vu et al. [116]	2009	Acad.	10	$[0.05, 0.5]$
Kodeswaran et al. [117]	2009	Acad.	15000	$[0.01, 2]$
McSherry et al. [94]	2010	Acad.	$16 \cdot 10^6$	$[0.1, 10]$
Bhaskar et al. [118]	2010	Acad.	990	$[0.12, 2.6]$
Pathak et al. [119]	2010	Acad.	48.9	$[0.01, 0.4]$
Machanavajjhala et al. [120]	2011	Acad.	103.7	$[0.5, 3]$
Chen et al. [121]	2011	Acad.	1210	$[0.5, 1.5]$
Cormode et al. [122]	2011	Acad.	1630	$[0.1, 1]$
Bonomi et al. [123]	2012	Acad.	150	$[0.01, 10]$
Narayan et al. [124]	2012	Acad.	32	0.69
Chen et al. [125]	2012	Acad.	1000	1
Ács et al. [89]	2012	Acad.	N/A	1
Xiao et al. [88]	2012	Acad.	49	$[0.05, 0.2]$
Li et al. [126]	2012	Acad.	15000	$[0.1, 2.5]$
Vinterbo et al. [127]	2012	Acad.	N/A	2.037
Mohammed et al. [128]	2012	Acad.	49	$[0.1, 1]$
Chaudhuri et al. [129]	2013	Acad.	494	$[0.01, 2]$
Uhlerop et al. [130]	2013	Acad.	40.84	$[0.1, 0.4]$
Andrés et al. [131]	2013	Acad.	10	$[0.01, 15]$
Wang et al. [132]	2013	Acad.	N/A	$[18.4, 4474]$
Backes et al. [133]	2013	Acad.	N/A	$[0.18, 0.5]$
Fredrikson et al. [21]	2014	Acad.	2.64	$[0.25, 100]$
Erlingsson et al. [32]	2014	Comm.	14000	25.63
Fan et al. [134]	2014	Acad.	990	$[0.01, 1]$
Lu et al. [135]	2014	Acad.	N/A	$[0.1, 1]$
Shen et al. [136]	2014	Acad.	229.9	$[1, 3]$
Xiao et al. [137]	2015	Acad.	6442.9	$[0.2, 1]$
Li et al. [138]	2015	Acad.	N/A	1
Han et al. [139]	2015	Acad.	40	$[7, 15]$
Phan et al. [140]	2016	Acad.	15.6	$[0.1, 6.4]$
Wang et al. [91]	2016	Acad.	1710.7	$[0.1, 1]$
Lin et al. [141]	2016	Acad.	4300	$[0.1, 0.6]$
Chen et al. [93]	2016	Acad.	1634.5	$[0.25, 1.25]$
Shen et al. [142]	2016	Acad.	1000.2	$[0.1, 1]$
Nguyen et al. [143]	2016	Acad.	9000	$[0.05, 0.8]$
Barbosa et al. [96]	2016	Acad.	4.46	$[0.001, 1.08]$
Qin et al. [144]	2017	Acad.	183.8	$[0.01, 7]$
Ding et al. [8]	2017	Comm.	3000	1.67
Yin et al. [145]	2017	Acad.	2	$[0.005, 15]$
Chen et al. [146]	2017	Acad.	48.8	$[0.1, 1]$
Eibl et al. [92]	2017	Acad.	14.05	$[0.1, 2]$
Bittau et al. [147]	2017	Acad.	10000	$[1.2, 2.25]$
Xiong et al. [148]	2018	Acad.	990	$[0.1, 1]$
Zhu et al. [149]	2018	Acad.	103.7	$[0.1, 1]$
Lyu et al. [150]	2018	Acad.	51	$[0.5, 1]$
Khavkin et al. [151]	2018	Acad.	1000	0.01
Fan et al. [152]	2018	Acad.	60	$[0.1, 1]$
Kim et al. [153]	2018	Acad.	1000	$[0.98, 3.63]$
Choi et al. [154]	2018	Acad.	5	$[0.5, 1]$
Asada et al. [155]	2019	Acad.	36001	$[0.001, 0.01]$
Wilson et al. [156]	2019	Acad.	N/A	$[0.25, 1]$
Zhao et al. [157]	2019	Acad.	989.8	$[1, 4]$
Sadilek et al. [69]	2019	Comm.	$300 \cdot 10^3$	0.66
Foote et al. [78]	2019	Govt.	372455	3
Bailie et al. [84]	2019	Govt.	23401.9	0.693
Fernandes et al. [158]	2019	Acad.	797	$[10, 30]$
Wang et al. [159]	2019	Acad.	546.4	$[0.5, 3]$
Steil et al. [160]	2019	Acad.	11.4	$[1, 70]$
Helmer et al. [54]	2019	Comm.	N/A	N/A
Cho et al. [161]	2020	Acad.	N/A	$[0.05, 2]$
Lee et al. [162]	2020	Acad.	1361	$[0.1, 1]$
Ou et al. [90]	2020	Acad.	N/A	$[0.1, 2]$
Chamikara et al. [163]	2020	Acad.	202.6	$[0.5, 8]$
Aktay et al. [11]	2020	Comm.	N/A	2.64
Bavadekar et al. [75]	2020	Comm.	2700	1.68
Microsoft Research [65]	2020	Comm.	$100 \cdot 10^3$	0.022
Messing et al. [64]	2020	Comm.	38000	1.453
Herdağdelen et al. [108]	2020	Comm.	6950	2
Paré et al. [18]	2020	Comm.	4.95	6.8
Wang et al. [164]	2020	Acad.	1200	$[1, 9]$
Sun et al. [165]	2020	Acad.	32.7	$[0.8, 20]$
Johnson et al. [72]	2020	Comm.	N/A	N/A
Lin et al. [166]	2021	Acad.	N/A	$[0.1, 1]$
Bavadekar et al. [10]	2021	Comm.	1500	2.19
Pereira et al. [17]	2021	Comm.	32.65	0.2
Kenny et al. [79]	2021	Govt.	$330 \cdot 10^3$	19.61
Apple Research [76]	2021	Comm.	N/A	8
Qian et al. [167]	2021	Acad.	27.7	$[0.27, 2.94]$
Yue et al. [168]	2021	Acad.	70.04	$[0.5, 20]$
Cunningham et al. [169]	2021	Acad.	33278	5
Google Research [67]	2021	Comm.	N/A	N/A
Adeleye et al. [60]	2022	Comm.	$325 \cdot 10^6$	0.998
Apple Research [7]	2022	Comm.	$100 \cdot 10^3$	$[2, 8]$
Han et al. [170]	2022	Acad.	573.7	$[1, 9]$
Rogers et al. [16]	2022	Comm.	$690 \cdot 10^3$	34.90
Rogers et al. [86]	2022	Comm.	50000	9.7
Spectus [74]	2022	Comm.	6800	10
Chen et al. [171]	2022	Acad.	62.5	$[1, 10]$
Gopi et al. [109]	2023	Comm.	17	12
Sadilek et al. [80]	2023	Govt.	$128 \cdot 10^3$	35.62
Abowd et al. [12]	2023	Govt.	$330 \cdot 10^3$	$[26.3, 34.3]$
Chen et al. [172]	2023	Acad.	N/A	$[0.1, 10]$
Sun et al. [173]	2023	Acad.	50	$[0.1, 2]$
Xu et al. [174]	2023	Acad.	1220	$[3.90, 9.67]$
Xu et al. [56]	2023	Acad.	3000	$[0.99, 6.82]$
Chen et al. [175]	2023	Acad.	60	N/A
Acharya et al. [176]	2023	Acad.	9000	$[2, 6]$
Hu et al. [177]	2023	Acad.	60	$[0.1, 16]$
Keeler et al. [178]	2023	Acad.	1000	$[0.025, 0.8]$
Wikimedia Foundation [61]	2023	Govt.	3.5	2
Wikimedia Foundation [62]	2023	Govt.	30	0.1
Apple Research [59]	2023	Comm.	4500	1
Bian et al. [70]	2024	Comm.	4000	2
Zhang et al. [57]	2024	Comm.	22280	1
Zhang et al. [57]	2024	Comm.	22280	2
Badrinarayanan et al. [87]	2024	Comm.	20000	4.5
Hod et al. [85]	2024	Govt.	165.9	9.98
Batool et al. [179]	2024	Acad.	1258	$[0.001, 2]$
Qashlan et al. [180]	2024	Acad.	$2 \cdot 10^6$	$[0.001, 10^3]$
Novado et al. [181]	2024	Acad.	1000	$[1, 8]$
Ju et al. [182]	2024	Acad.	10.4	$[1, 5]$
Shanmugarasa et al. [183]	2024	Acad.	150	$[1, 20]$
Agrawal et al. [184]	2024	Acad.	1565.9	$[1, 1000]$
Sun et al. [105]	2024	Acad.	1000	10
Xu et al. [107]	2024	Comm.	N/A	$[6, 13]$