electronics

MDPI

Technical Note

Group Privacy: An Underrated but Worth Studying Research Problem in the Era of Artificial Intelligence and Big Data

Abdul Majeed [1,*], Safiullah Khan [2] and Seong Oun Hwang [1,*]

[1] Department of Computer Engineering, Gachon University, Seongnam 13120, Korea
[2] Department of IT Convergence Engineering, Gachon University, Seongnam 13120, Korea; safi@gachon.ac.kr
* Correspondence: ab09@gachon.ac.kr (A.M.); sohwang@gachon.ac.kr (S.O.H.); Tel.: +82-31-750-5327 (S.O.H.)

Abstract: *Introduction:* Recently, the tendency of artificial intelligence (AI) and big data use/applications has been rapidly expanding across the globe, improving people's lifestyles with data-driven services (i.e., recommendations, smart healthcare, etc.). The synergy between AI and big data has become imperative considering the drastic growth in personal data stemming from diverse sources (cloud computing, IoT, social networks, etc.). However, when data meet AI at some central place, it invites unimaginable privacy issues, and one of those issues is group privacy. Despite being the most significant problem, group privacy has not yet received the attention of the research community it is due. *Problem Statement:* We study how to preserve the privacy of particular groups (a community of people with some common attributes/properties) rather than an individual in personal data handling (i.e., sharing, aggregating, and/or performing analytics, etc.), especially when we talk about groups purposely made by two or more people (with clear group identifying markers), for whom we need to protect their privacy as a group. *Aims/Objectives:* With this technical letter, our aim is to introduce a new dimension of privacy (e.g., group privacy) from technical perspectives to the research community. The main objective is to advocate the possibility of group privacy breaches when big data meet AI in real-world scenarios. *Methodology:* We set a hypothesis that group privacy (extracting group-level information) is a genuine problem, and can likely occur when AI-based techniques meet high dimensional and large-scale datasets. To prove our hypothesis, we conducted a substantial number of experiments on two real-world benchmark datasets using AI techniques. Based on the experimental analysis, we found that the likelihood of privacy breaches occurring at the group level by using AI techniques is very high when data are sufficiently large. Apart from that, we tested the parameter effect of AI techniques and found that some parameters' combinations can help to extract more and fine-grained data about groups. *Findings:* Based on experimental analysis, we found that vulnerability of group privacy can likely increase with the data size and capacity of the AI method. We found that some attributes of people can act as catalysts in compromising group privacy. We suggest that group privacy should also be given due attention as individual privacy is, and robust tools are imperative to restrict implications (i.e., biased decision making, denial of accommodation, hate speech, etc.) of group privacy. *Significance of results:* The obtained results are the first step towards responsible data science, and can pave the way to understanding the phenomenon of group privacy. Furthermore, the results contribute towards the protection of motives/goals/practices of minor communities in any society. *Concluding statement:* Due to the significant rise in digitation, privacy issues are mutating themselves. Hence, it is vital to quickly pinpoint emerging privacy threats and suggest practical remedies for them in order to mitigate their consequences on human beings.

Keywords: group privacy; artificial intelligence; big data; analytics; privacy-preserving data publishing; utility; data mining; differential privacy; clustering; social network

Citation: Majeed, A.; Khan, S.; Hwang, S.O. Group Privacy: An Underrated but Worth Studying Research Problem in the Era of Artificial Intelligence and Big Data. *Electronics* **2022**, *11*, 1449. https://doi.org/10.3390/electronics11091449

Academic Editor: Qingqi Pei

Received: 6 April 2022
Accepted: 27 April 2022
Published: 30 April 2022

1. Introduction

Data owners such as hospitals, insurance companies, and banks collect huge amounts of data on a daily basis. The sole purpose of the data collection is to improve the quality of

service as well as to serve customers in a seamless manner. With the rapid development of pervasive computing, data collection has become easier. With more digital tools, a huge amount of data including intimate details of our life (for example, *our demographic characteristics, our social status, where we live or work, what we buy, when we buy, what type of car we own, where we go on weekends, what type of mobile we use, what type of websites we search, our profession, our monthly earning, our music choices, our dress choices, our religion, our political views, illnesses we may have or have had, our hobbies)* can now easily be collected. With such detailed data, many data owners can construct detailed profiles about us that can be used for healthcare/product recommendations. Although these data have a lot of potential in influencing science and societies, privacy issues can limit its use due to its processing in a black-box manner [1,2]. Although the privacy domain was investigated well from different perspectives, its landscape is changing continuously amid technical developments [3]. In recent years, privacy has become one of the most researched topics, and many developments are originating from all parts of the world to address this social issue [4,5].

1.1. The Emergence of Group Privacy Issues: A New Dimension

In the early days of the COVID-19 pandemic, digital tools were mostly used to curb the spread by tracing potentially infected individuals who had been in close contact with infected individuals [6,7]. In these tools, a variety of personal information data are used/collected to identify the probably infected individuals as quickly as possible.

Although COVID-19 can infect anyone across the globe, its emergence in societies/communities/groups that are already facing some discrimination/backlash from society due to their controversial behaviors or activities can be severe, leading to many types of harm. We refer to these communities as groups and preserving their privacy is vital to give them sufficient protection and respect in society. We demonstrate two real-world examples in which group privacy breaches have caused severe consequences in Figure 1. As shown in Figure 1, group privacy issues can cause more consequences compared to individual privacy issues. Amid the intrusion into one's personal life in the above two real-world examples, some members committed suicide due to aggressive technologies' usage of big data and AI technologies with the fine-grained personal data of targeted groups [8–10]. Due to technical advancements in most fields, privacy threats to group privacy are increasing in recent times, and quick remedies are needed towards this genuine problem of big data technologies.

1.2. Comparative Analysis of Individual and Group Privacy in the Era of Big Data and AI

After a detailed analysis of the published literature and existing developments, we present a classification of individual and group privacy threats in Figure 2.

In Figure 2, we classify privacy threats into three categories based on the time scale. As shown in Figure 2, group privacy is likely to be one of the main threats faced by individuals in the era of AI and big data [11]. The main reason for the increase in group privacy is the increasing benefits of data analysis for accomplishing multiple goals (e.g., pandemic control, effective decision making, etc.) using data-driven approaches [12].

The major contributions of this article are summarized as follows: (i) It provides an overview of group privacy which is an urgent problem to be addressed in the context of big data and AI technologies to lower privacy breaches and the corresponding harms to society. Specifically, it highlights the emergence, needs, and transition from individual to group privacy that can likely be a major threat in the information privacy area in the near future; (ii) It identifies three main research tracks of individual privacy preservation and summarizes state-of-the-art developments in each track; (iii) It discusses existing approaches that have been devised for group privacy preservation, and pertinent threats to group privacy in the era of AI and big data that remained unexplored in the recent literature; (iv) It provides a case study using two real-world benchmark datasets highlighting the privacy issues that can emanate from it based on the values of the attributes; (v) It highlights

various kinds of group privacy problems in different computing paradigms (i.e., cloud computing, social networks, Internet of Things, location-based systems, etc.) that have not been covered in prior research; (vi) It lists potential research directions in the area of AI and big data for group privacy preservation that needs more research/development from both academics and the industry; (vii) To the best of our knowledge, this is the first detailed work on group privacy, and is a timely contribution toward responsible data science amid continuous technological advancements.

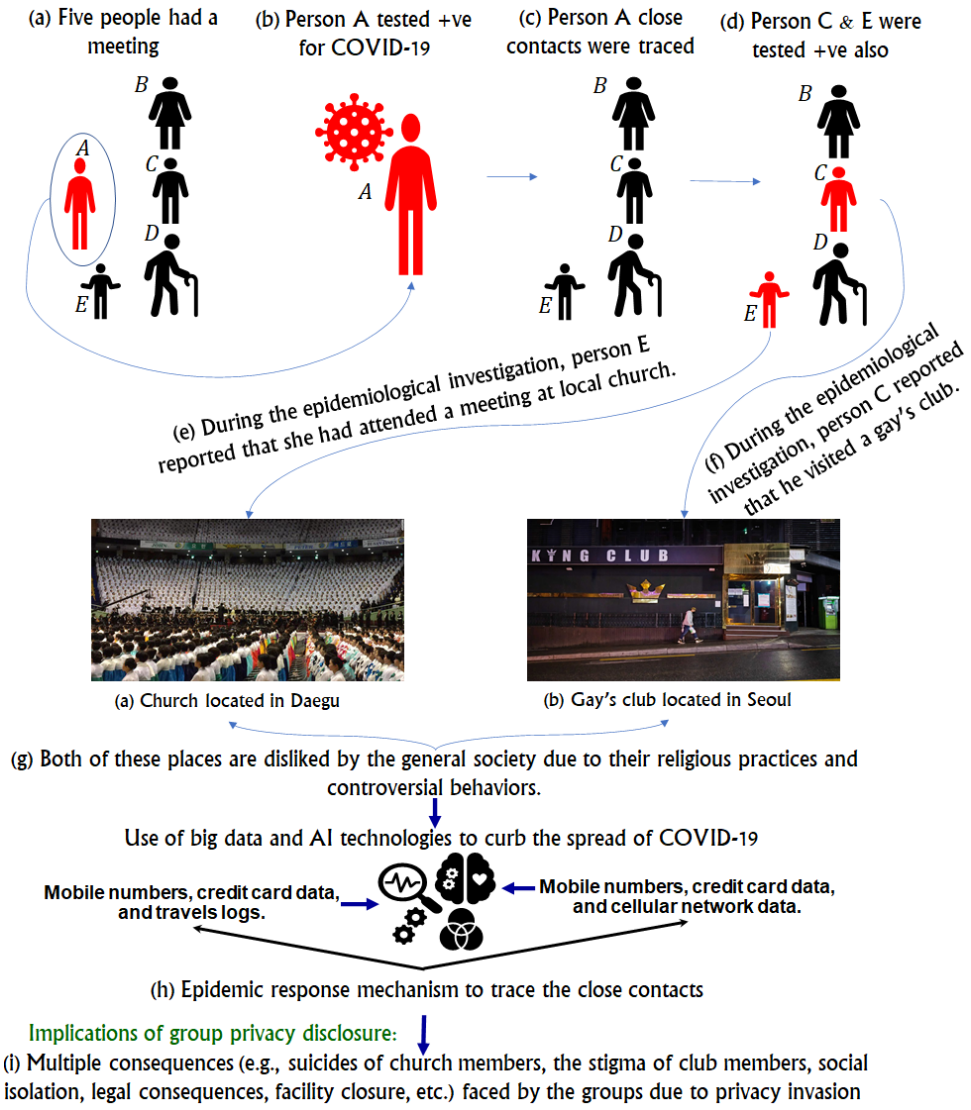


Figure 1. Overview of the emergence of group privacy issues in real-world cases.

The rest of this article is organized as follows. Section 2 presents information privacy concept, and three main research tracks for individual privacy preservation. Section 3 discusses the group privacy concept, threats, and the recent developments with regard to group privacy preservation. Section 4 presents a case study to show the significance of group privacy using two real-world datasets. Section 5 presents the future research outlook of the privacy domain in the era of big data and AI and lists various research directions that are vital to combat group privacy issues. Finally, we conclude this paper in Section 6.

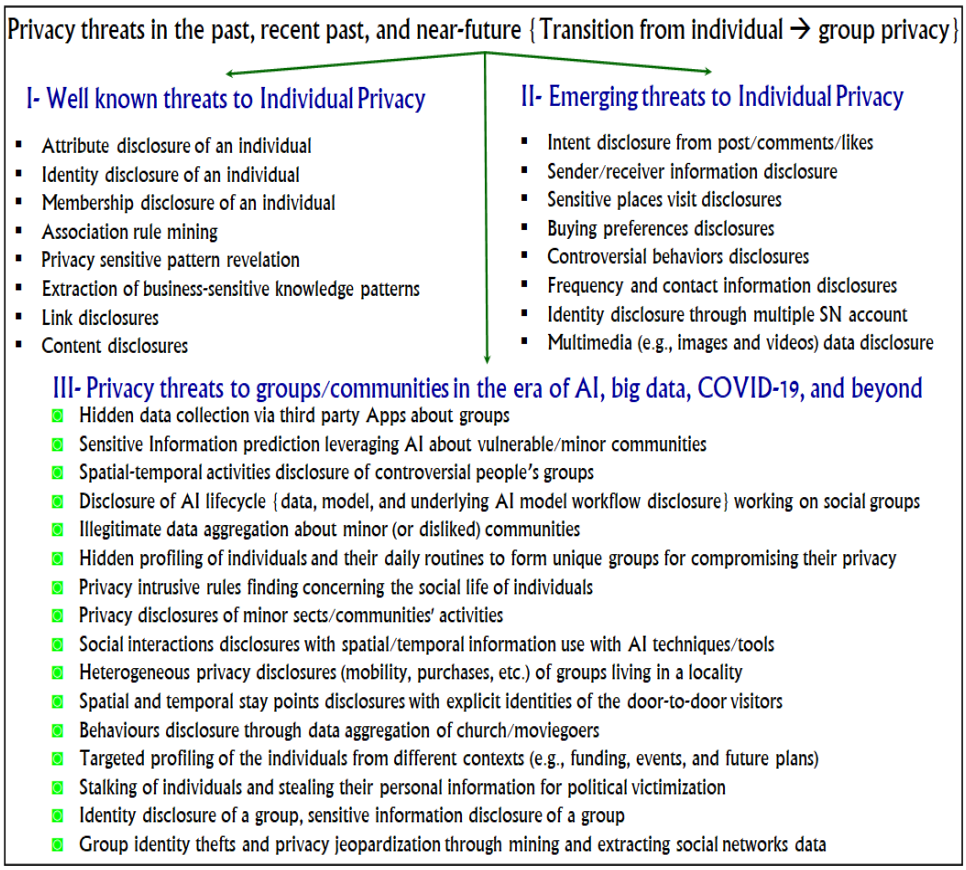


Figure 2. Classification of privacy threats (transition from individual → group privacy in the near-future).

2. State-of-the-Art Privacy Preserving Approaches

In the information privacy domain, personal data can be represented with the help of either tables or graphs depending on the data owners. For example, hospitals usually collect and process personnel in tabular form. In contrast, social network (SN) data are mostly represented with the help of graphs. In Figure 3, we provide an overview of the ten most widely used data representation types along with the corresponding data owners. In this work, we assume personal data encompassed in the tables and graph, respectively.

To overcome privacy issues, five main techniques were applied to personal data, as shown in Figure 4. The selection of the technique depends on the nature of data, computing environment, and desired goals. Each technique has certain benefits over one another, either in terms of preserving privacy or computing resources. For example, encryption-based methods are usually slower than anonymization approaches [13].

These approaches employ variety of operations such as generalization [14], suppression [15], bucketization [16], hash functions [17], cryptographic primitives [18], lattice-based encryption [19], parameter sharing [20], masking [21,22], pseudonyms [23–26], and joint operations [27–30] in order to preserve the privacy of the individual. Recently, machine learning (ML) approaches have also been employed to preserve the privacy of individuals in data analysis and publishing [31–35]. ML approaches have significantly improved the traditional privacy preserving approaches by extracting attribute level information from data. Furthermore, ML approaches have created synergy with most of the approaches listed in Figure 4 to effectively preserve individual privacy [36–39].

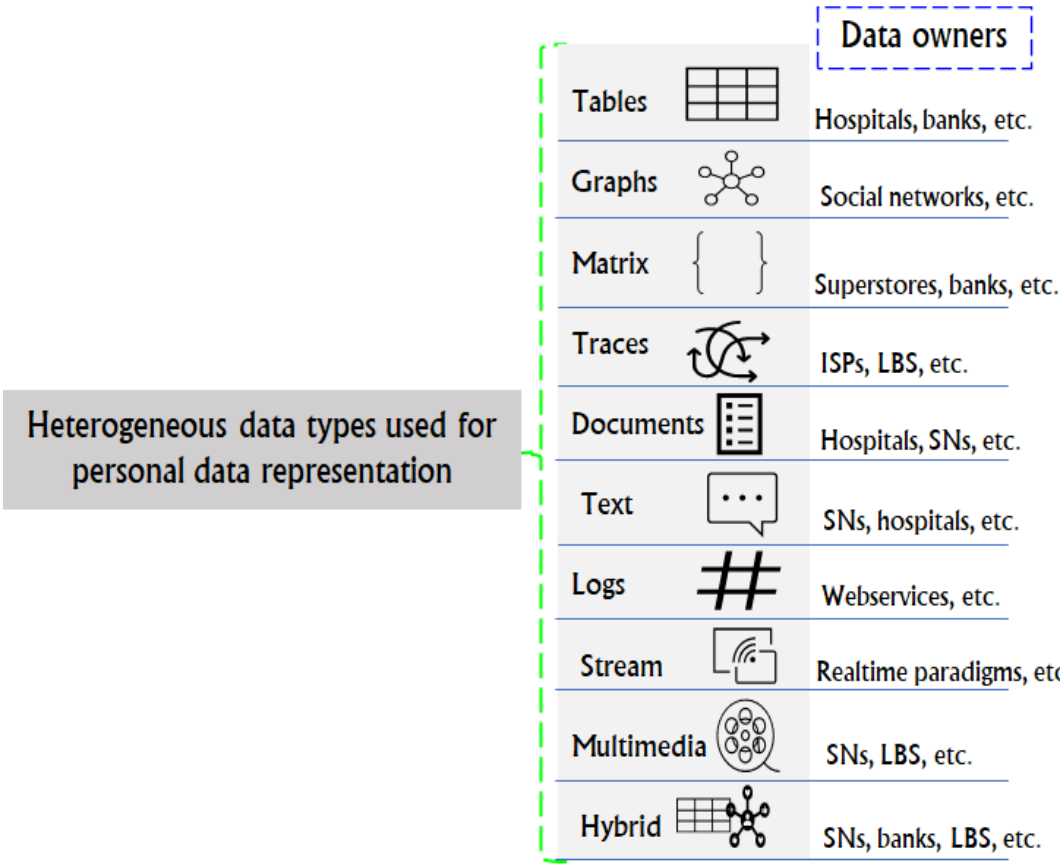


Figure 3. Overview of the heterogeneous data types used to represent personal data.

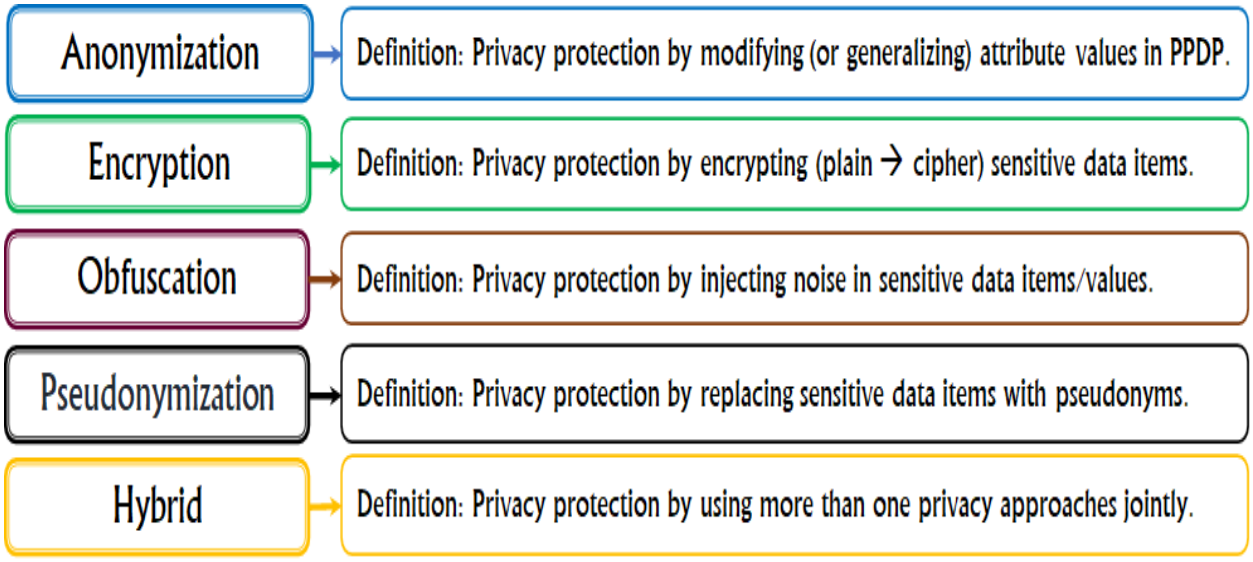


Figure 4. Practical approaches for the privacy preservation of personal data.

2.1. Research Tracks for Individual Privacy Preservation

Thus far, considerable algorithms and prototypes have been developed for privacy preservation. These approaches were applied to the tabular data and were extended to other styles of data. We classify the existing developments into three potential research tracks such as track *A*, track *B*, and track *C* (as shown in Figure 5).

2.1.1. Individual Privacy Preservation and SOTA Approaches in Track A

The research in track A has been underway since 2002 with the Sweeney study named 'simple demographics often identify people uniquely' [40]. According to her findings, the identification of unique people is possible at extremely higher percentages based on the following three combinations of demographic values:

- Zip code (five-digits), gender, date of birth → **87%**
- Place of residence, gender, date of birth → **50%**
- Country of origin, gender, date of birth → **18%**

The remarkable developments in track A are k-anonymity [41], ℓ-diversity [42], t-closeness [43], and their improved versions [44–51]. An overview of the k-anonymity model is given in Figure 6.

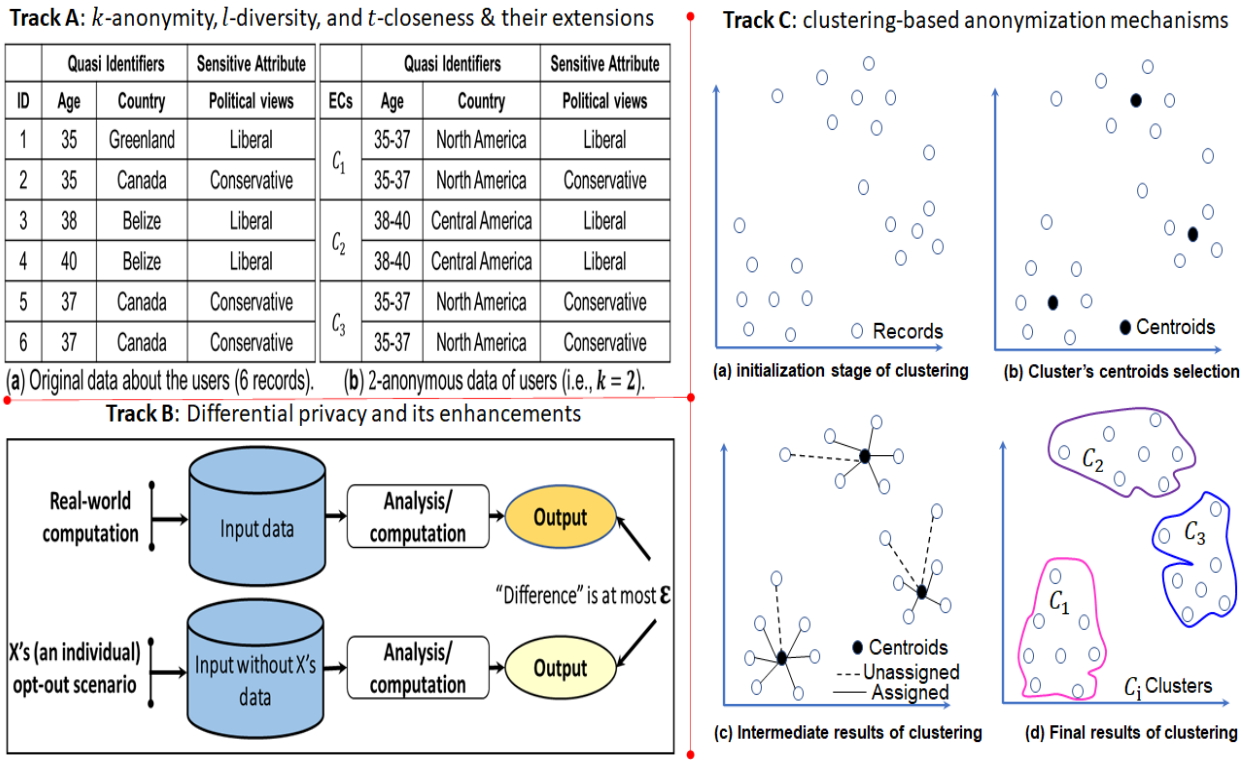


Figure 5. Overview of three main research tracks in individual privacy preservation.

According to this model, the re-identification ability of a person X is $\frac{1}{k}$ in published data. Due to the conceptual simplicity and first approach for privacy preservation, this model was extensively studied and improved from multiple perspectives even in the recent literature. These three approaches (e.g., k-anonymity [41], ℓ-diversity [42], and t-closeness) were extensively investigated regarding privacy preservation from different contexts. In addition, in most formats of data, the k-anonymity model was extensively applied to ensure some form of privacy. In Figure 7, we present an overview of the ℓ diversity model that is an extended version of k-anonymity.

(a) Original data table to be anonymized

Quasi Identifiers (QIs)				SA Info
Education	**Race**	**Sex**	**Age**	**Salary**
Bachelors	White	M	39	> 50K
Bachelors	White	M	50	≤ 50K
HS-grad	White	M	38	≤ 50K
11th	Black	M	53	> 50K
Bachelors	Black	F	28	≤ 50K
Masters	White	F	37	> 50K
9th	Black	F	49	≤ 50K
HS-grad	White	F	52	> 50K
Masters	White	F	31	≤ 50K
Bachelors	White	M	42	> 50K

(b) Original data after being anonymized with $k = 2$

Quasi Identifiers (QIs)				SA Info
Education	**Race**	**Sex**	**Age**	**Salary**
Bachelors	White	M	39-42	> 50K
Bachelors	White	M	39-42	> 50K
*	White	M	50-52	≤ 50K
*	White	M	50-52	> 50K
*	White	*	37-38	≤ 50K
*	White	*	37-38	> 50K
High	*	F	28-31	≤ 50K
High	*	F	28-31	≤ 50K
Low	Black	*	49-53	> 50K
Low	Black	*	49-53	≤ 50K

Figure 6. Overview of the k-anonymity model used for the privacy preservation of the individual.

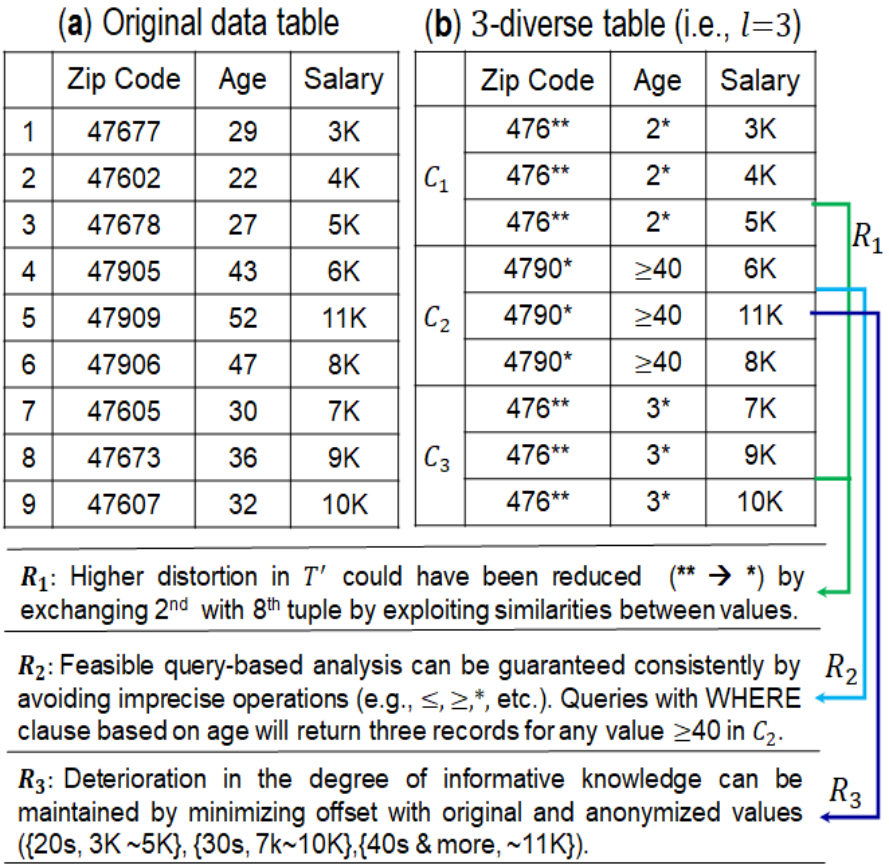


Figure 7. Overview of possible refinements in ℓ-diversity anonymity model.

According to this model, the probability of inferring the SA of an individual from data is $\frac{1}{\ell}$. Although ℓ-diversity helps in privacy preservation, anonymized data quality can be

lower due to the enforcement of hard constraints (e.g., ℓ). We present three refinements (R_1, R_2, R_3) to ℓ-diversity in Figure 7 that can be helpful in augmenting information availability in data.

The SOTA approaches published in the past five years regarding privacy preservation in track *A* are summarized as follows. Xu et al. [52] discussed privacy issues that can emanate from trajectory data publishing. Tu et al. [53] proposed a mechanism for protecting the privacy of individuals in trajectory data by jointly using *k*-anonymity, ℓ-diversity, and *t*-closeness concepts. Eom et al. [54] developed a privacy and utility preserving anonymization method based on surrogate vectors. The proposed method preserves individual privacy as well as satisfies ϵ-DP. Cao et al. [55] proposed a method based on the *k*-anonymity concept for both the location (i.e., geo-indistinguishability) and spatiotemporal event privacy preservation. Shaham et al. [56] developed a privacy-preserving mechanism based on the machine learning (ML) concepts. The proposed ML-based anonymization (MLA) framework preserved better utility and privacy in publishing location data. Cabrero et al. [57] developed a privacy preservation concept for deep learning (DL) models. The proposed PPDL concept yields superior privacy and utility results in DL paradigms. Guan et al. [58] devised a practical privacy-preserving approach for optimizing privacy-utility trade-off using DP-based clustering scheme named EDPDCS. The proposed EDPDCS method can yield an effective resolution of privacy and utility trade-offs in big data environments. Ashkouti et al. [59] proposed a new model based on ℓ-diversity concept for big data privacy preservation. The proposed method uses the concept of parallel and distributed computing in order to overcome the latency issues in large-scale data anonymization. Wang et al. [60] proposed a new anonymization method for publishing data containing multiple SA about individuals. The proposed method effectively preserves data utility and privacy and is based on the *t*-closeness concept. Mehta et al. [61] proposed an improved ℓ diversity model for privacy preservation in data publishing. The proposed model makes use of the MapReduce paradigm to anonymize big data. The proposed method has the ability to lower the information loss as well as the complications of the clustering process. Bazai et al. [62] proposed a subtree-based anonymization method with a highly efficient generalization strategy. The proposed method yields superior results in privacy and utility than SOTA anonymization techniques. Zouinina et al. [63] discussed a new anonymization technique based on multi-view micro aggregation. The proposed technique is based on the *k* anonymity concept and has many benefits in preserving structural utility as well better privacy preservation in PPDP. Recently, a new anonymization approach based on the bucketization concept was given by Jayapradha et al. [64]. The proposed approach makes use of the *k*-anonymity and slicing concepts in order to preserve the privacy of data encompassing multiple SAs about individuals. The proposed approach has the ability to provide a solid defense against five types of privacy attacks such as background knowledge attack, fingerprint correlation attack, membership attack, quasi-identifiers attack, and non-membership attack. Ito et al. [65] developed a new anonymization method for preserving the privacy of individuals in transactional data. The proposed method assists in selecting the optimized value of *k*, and is applicable in a wide range of data-driven applications. All approaches cited above have assisted in effectively preserving individual person privacy in heterogeneous domains/applications.

2.1.2. Individual Privacy Preservation and SOTA Approaches in Track *B*

The research in track *B* started in 2006 with the Dwork study/concept named 'Differential privacy'. Since its inception in 2006, it has been rigorously investigated from multiple perspectives as well as applications. The remarkable development in track *B* is differential privacy (DP) [66] and its improved versions [67–72]. Since its inception in 2006, DP has been extensively studied in the literature and has become a benchmark for privacy preservation in data analysis. A conceptual overview of the DP model is shown in Figure 8. DP from an attacker's viewpoint can be defined as DP is safeguarding the leakage of information in a way that only yields noisy random/aggregated information about an individual from any dataset.

For a dataset, D, of n individuals, this implies that DP will not expose the information/data of the last individual if an adversary holds the information about $n-1$ individuals. Due to such robust mathematical privacy guarantees, DP is one of the most studied concepts in the information privacy domain. It uses the noise addition and randomization operation in order to add more confusion in the output of statistical queries. Furthermore, it yields different answers to the same but repeated queries to hide the structural properties of the datasets from the attackers.

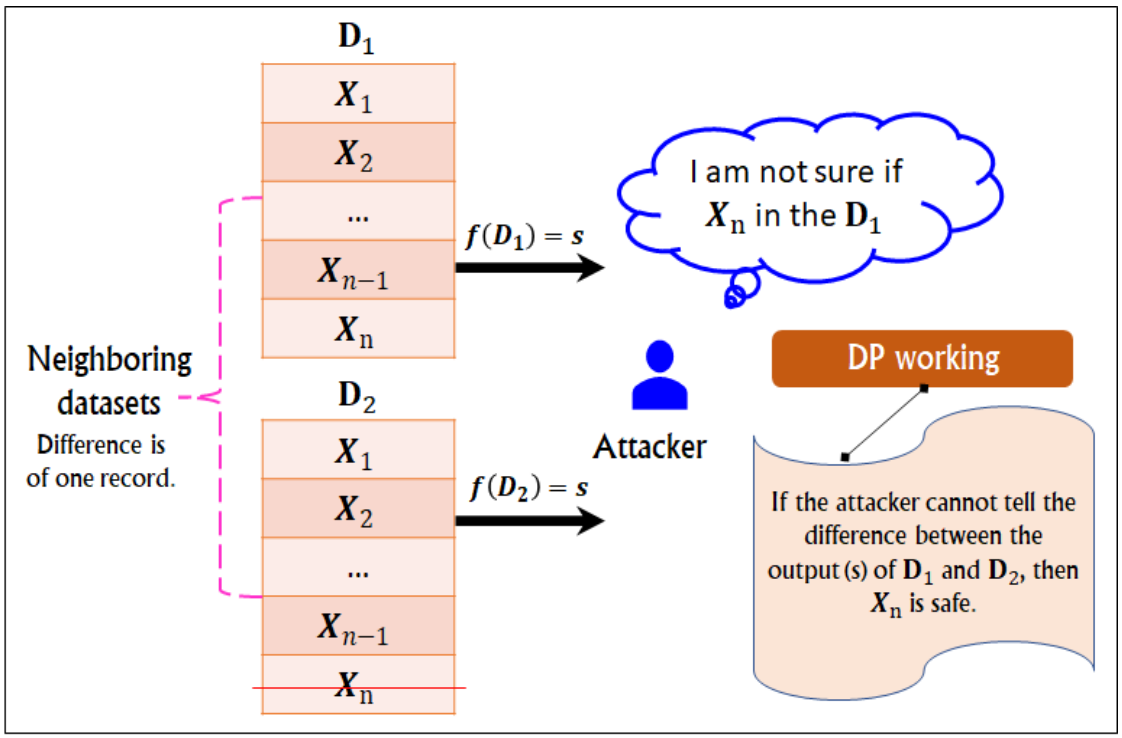


Figure 8. Overview of differential privacy model.

Under DP, an anonymity algorithm, A, satisfies ϵ-DP if for all subsets $S \subseteq Range(A)$ and for all T_1, T_2 such that $d(T_1, T_2) = 1$ (e.g., T_1 differs from the T_2 by just one tuple):

$$\frac{Pr(A(D_1) \in S)}{Pr(A(D_2) \in S)} \leq exp(\epsilon) \tag{1}$$

where ϵ denotes the privacy loss budget, and its value is generally higher than 0 (i.e., $\epsilon > 0$).

Later, many enhancements of the DP were suggested to increase its performance by injecting a relatively small amount of noise, the relaxation of the hard constraints/parameters, and lower/upper bounds for the DP. A popular version of the DP is (ϵ,δ)-DP, which is mathematically written as follows:

$$Pr(A(D_1) \in S) \leq exp(\epsilon) Pr(A(D_2) \in S) + \delta \tag{2}$$

In Equation (2), δ denotes the degree of relaxation, and its value is $\delta \in [0, 1]$. However, if $\delta = 0$, then A obeys traditional ϵ-DP guarantees. The selection of a suitable value for δ is very challenging, and was set to $10^{-7} \leq \delta \leq 10^{-10}$ in recent applications. For query responses, R, the probability (Pr) of the DP model is expressed in Equation (3):

$$\frac{Pr(A(D_1) = R)}{Pr(A(D_2) = R)} \leq e^{\epsilon} \tag{3}$$

In recent years, the DP concept has been expanded to various domains, such as social networks (SNs), Internet of Medical Things (IoMT), Internet of Things (IoT), and textual

data. It has two famous settings as local and global. In SN data, its two popular versions are nodes and edge DP.

The SOTA approaches published in the past five years regarding privacy preservation in track *B* are summarized as follows. Cai et al. [73] proposed a DP-based privacy-preserving model for big data trading in IoT systems. The proposed method approximates range counting and improvises data utility for legitimate information consumers. Zheng et al. [74] developed an *epsilon*-DP-based method for data sharing in industrial IoT environments with strong privacy guarantees. The proposed method yields superior results to existing SOTA approaches in data sharing towards Industry 4.0. Huo et al. [75] discussed a privacy protection model for IoT environments based on DP principles. The proposed model preserves the location privacy, data privacy, identity, and query privacy. Furthermore, the proposed model has abilities to work with Industry 4.0 technologies. Bagdasaryan et al. [76] analyzed the effect of the DP model in the ML environments and concluded that the accuracy of DP-SGD drops significantly if higher ϵ is used in neural network training. Wang et al. [77] proposed a local DP (LDP) model for mining patterns from set-valued data. The proposed model has the ability to answer sensitive questions via queries while preserving users' privacy. Li et al. [78] developed a DP-based model for privacy preservation in image data. The proposed model is the first practical approach that shows that facial privacy is measurable. Iwendi et al. [79] developed the first practical solution for privacy protection in unstructured medical datasets. The proposed approach makes use of the DP combined with negated assertions to improve privacy in medical domains. The experimental analysis indicates a significant improvement in the privacy and utility trade-off compared to existing methods. Nautsch et al. [80] developed a practical privacy-preserving approach for speech recordings using DP concepts. The proposed method also suggested many metrics for accurately evaluating the privacy of speech data. Sharma et al. [81] developed a DP-based method for privacy-preserving data analytics. The proposed method has higher significance in the healthcare information system. Ye et al. [82] developed an LDP-based privacy-preserving approach named perturbation calibration for key-value data. The proposed approach helps in frequency and mean estimation from large and high-dimensional data. Finally, due to robust privacy guarantees, DP has been extensively used in AI environments to preserve the privacy and utility of individuals [83–87]. In the coming years, DP will be an integral part of many emerging technologies with regard to privacy preservation [88]. Furthermore, it is one of the most widely used techniques in the cloud, edge, and fog computing environments for privacy preservation against active attackers [89–91]. Furthermore, DP adoption in IoT environments is significantly higher than it is in its counterparts [92,93].

In recent years, DP has helped secure AI models from malevolent adversaries. In this regard, Arachchige et al. [94] developed an LDP-based privacy-preserving solution for deep learning (DL). The proposed method introduced three modules to preserve privacy in the training of a convolutional neural network (CNN). Chamikara et al. [95] developed a distributed perturbation algorithm (DPA) to preserve the privacy of ML algorithms. Through extensive experimental analysis, the proposed method was considered an excellent solution for preserving privacy in distributed environments. Abramson et al. [96] outlined a prototype for privacy preservation for the distributed learning paradigms. The proposed prototype was applied to mental health care data for performance verification. Thapa et al. [97] discussed many promising applications of the DP in the FL and split learning domains. Through code implementation, the authors verified the DP potentials in these two domains regarding privacy preservation. Wang et al. [98] extended the (ϵ,δ) DP use in data collection scenarios for ML algorithms. The proposed method significantly outperformed SOTA studies while preserving both utility and privacy in numeric/categorical data. In recent years, SOTA approaches that can enhance the privacy of AL (ML + DL) models are increasing at a rapid pace [99–101]. Rahali et al. [102] developed a DP-based approach for recommendation systems with an optimized utility–privacy trade-off. The proposed

approach can yield better preserver privacy and utility, and it is resilient against averaging attacks in recommendation systems.

2.1.3. Individual Privacy Preservation and SOTA Approaches in Track *C*

The clustering methods (e.g., *k* means, *k* medoids, *k* means ++, hierarchical, partitional, DBSCAN, etc.) have improved the traditional anonymization approaches from multiple perspectives. The synergy of the clustering methods with the anonymization mechanisms has particularly improved the utility aspects of the PPDP process. These techniques group the records based on similarities to ensure higher privacy and utility. We demonstrate an overview of the clustering-based anonymization in Figure 9. In Figure 9a, original data to be sanitized are shown. In Figure 9b, clustering results are shown in which users based on the similarities are clustered, and the corresponding anonymized data are shown in Figure 9c.

(a) Original data table to be anonymized

Quasi Identifiers (QIs)				SA Info
Education	**Race**	**Sex**	**Age**	**Salary**
Bachelors	White	M	39	> 50K
Bachelors	White	M	50	≤ 50K
HS-grad	White	M	38	≤ 50K
11th	Black	M	53	> 50K
Bachelors	Black	F	28	≤ 50K
Masters	White	F	37	> 50K
9th	Black	F	49	≤ 50K
HS-grad	White	F	52	> 50K
Masters	White	F	31	≤ 50K
Bachelors	White	M	42	> 50K

(b) Original data after clustering (a.k.a grouping) process

Quasi Identifiers (QIs)				SA Info
Education	**Race**	**Sex**	**Age**	**Salary**
Bachelors	White	M	39	> 50K
Bachelors	White	M	42	> 50K
Bachelors	White	M	50	≤ 50K
HS-grad	White	M	52	> 50K
HS-grad	White	M	38	≤ 50K
Masters	White	F	37	> 50K
Masters	White	F	31	≤ 50K
Bachelors	Black	F	28	≤ 50K
11th	Black	M	53	> 50K
10th	Black	F	49	≤ 50K

(c) Original data after being anonymized for k=2

Quasi Identifiers (QIs)				SA Info
Education	**Race**	**Sex**	**Age**	**Salary**
Bachelors	White	M	39-42	> 50K
Bachelors	White	M	39-42	> 50K
*	White	M	50-52	≤ 50K
*	White	M	50-52	> 50K
*	White	*	37-38	≤ 50K
*	White	*	37-38	> 50K
High	*	F	28-31	≤ 50K
High	*	F	28-31	≤ 50K
Low	Black	*	49-53	> 50K
Low	Black	*	49-53	≤ 50K

Figure 9. Practical example of personal data anonymization using clustering techniques for achieving 2-anonymity (e.g., $k = 2$).

The major developments in track *C* are *k* means, *k* medoids, hierarchal, DBSCAN, and partitional clustering-based randomization techniques. These techniques have revolutionized the privacy domain by improving various technical aspects of the anonymization methods stated in track *A*. The remarkable algorithms in this track are *k*-means++-based anonymity [103], *k*-members-based anonymity [104], *k* means for ℓ diversity [105], and *K*-medoids-based anonymization [106]. These approaches have improved various aspects of traditional anonymization methods by loosening the strict parameters. These approaches yield better results in terms of both privacy and utility in data publishing. Furthermore, this technique was applied to heterogeneous data styles with slight modifications [107]. Guo et al. [108] proposed a fast anonymization technique based on the clustering concept. The proposed technique works well on the stream data and satisfies the ℓ-diversity property. Onesimu et al. [109] developed an anonymization method based on the clustering concept for IoT scenarios. The proposed method uses a modified *k*-means clustering scheme to achieve *k*-anonymity properties. Sopaoglu et al. [110] proposed a novel method for stream data anonymization. The proposed method fulfills the *k*-anonymity property by taking user's privacy preferences as a parameter from the user. The proposed method effectively satisfies the privacy and utility trade-off in the PPDP. Yang et al. [111] developed a new clustering-based anonymization method for stream data anonymization. The proposed method can anonymize the incomplete stream data with a better balance of utility and privacy. Nasab et al. [112] developed a computationally efficient framework for large-scale stream data anonymization based on *k*-anonymity concepts. The proposed framework can strike the balance well between utility and privacy. Tekli et al. [113] proposed a new anonymization approach for transactional data based on (k, ℓ)-clustering. The proposed approach can yield higher utility than existing methods. Parameshwarappa et al. [114] proposed the clustering-based anonymization of sequential data. Guo et al. [115] proposed a

new clustering-based anonymization method in order to optimize the utility and efficiency trade-off. The proposed method extracts natural equivalent classes in order to lower the complications of the clustering process. Zheng et al. [116] developed an anonymization algorithm based on improved clustering. The proposed algorithm lowers the information loss by 20% on benchmark datasets. Siddula et al. [117] developed a new anonymization algorithm based on clustering concepts for SN data. The proposed algorithm ensures both k and ℓ properties in the graphs while anonymizing data. Zhao et al. [118] developed a new anonymity method based on clustering and DP concepts to overcome privacy issues in the trajectory data. Liu et al. [119] proposed a new privacy protection method for trajectory data based on k-means clustering. The proposed method satisfies the DP properties as well. Yan et al. [120] developed a clustering-based anonymization method for securing smart meter data. The proposed method utilizes the clustering concept jointly with the DP in order to preserve the privacy of smart meter data. Lan et al. [121] developed a novel anonymization method for skyline queries. The proposed method was experimentally tested on synthetic and real data. Along this line of research (e.g., clustering-based anonymization), many approaches have recently been developed to secure personal data as well permitting the performance of analytics on them [122–127].

In addition to the three generic tracks discussed above, privacy approaches can have different tracks in each data type. For example, privacy-preserving approaches in social networks data have five different tracks of research, as shown in Figure 10. These tracks are graph modification, DP-based graph data anonymization, privacy-aware graph computation, graph clustering, and hybrid anonymization approaches. More information about these categories can be gathered from the recent literature [128–133]. Similarly, in tabular data, the tracks can be classified into two; multiple quasi-identifiers and one SA, and multiple quasi-identifiers and two SA. In trajectory data, all above three tracks have been rigorously studied in the current literature. In conclusion, there exist plenty of methods for preserving individual privacy in different data types (e.g., tables, matrices, sets, logs, traces, images, streams, videos, text, documents, etc.).

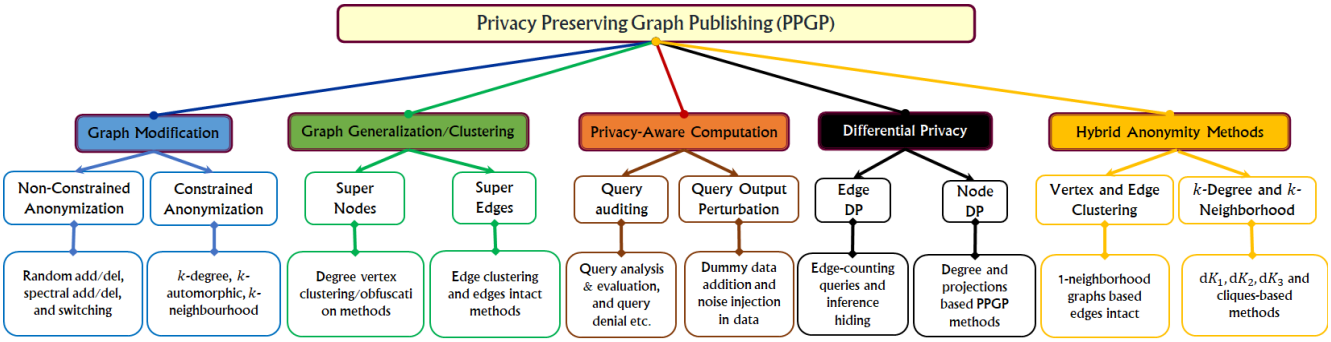


Figure 10. Research tracks in privacy preservation of social network data.

3. Group Privacy: A New Dimension of Privacy

In 2017, Taylor et al. [134] discussed the epistemological phenomenon of 'group privacy' in the big data analytics era for the first time. The authors discussed many useful concepts about group privacy from multiple perspectives (i.e., legal, technical, ethical, etc.). The group can refer to a number of individuals or things, and privacy is also about keeping personal information away from prying eyes. In simple words, group privacy is to preserve the privacy of a number of individuals' characteristics and habits. *With big data analysis, an individual's characteristics and habits can increasingly be taken to represent a cluster/class of similar individuals and, on their own, suffice to draw conclusions about a group.* In recent years, big data analytics can help in identifying groups from large and high dimensional data with ease [135]. Although analyzing groups can help in recommendation purposes, the revelation of group characteristics in terms of political affiliation, religious views, and controversial behaviors can have a range of negative consequences as stated

in the introduction section. Loi et al. [136] discussed two concepts about group privacy such as Type-a (i.e., interaction history or common goals) groups and Type-b (i.e., common features) groups that were not comprehensively discussed in the previous literature. In the recent literature, the need for individuals, groups, and even societies as a whole have also been stressed. Interestingly, some groups can be susceptible to privacy breaches/violations because of their class, race, gender or religion, sexual identity, or other intersectional characteristics/circumstances [137]. Recently, due to rapid advances in machine and deep learning techniques, groups can be algorithmically determined which can lead to biased decision-making about certain controversial groups. Hence, studying group privacy from theoretical, technical, and ethical perspectives has become more urgent than ever [138].

To emphasize the need for investigating/studying group privacy, we present a practical attack in location-based systems (LBSs) (https://elf11.github.io/2017/05/06/lbs-part-1.html, Accessed on: 16 March 2022) to infer group movement information in Figure 11. Let us consider that ACME is an insurance firm with a huge number of clients; the list of clients is one of the valuable business assets of the company and must be kept private. The employees of ACME visit their clients frequently. Before starting their trip, they utilized an LBS (i.e., Google location maps) which determined the optimized routes. Due to traffic congestion, the suggested routes can change frequently. Now, if the LBS is untrustworthy, it can reconstruct the entire client list with very high probability by orchestrating frequent route queries that emerge from ACME. To prevent this happening (e.g., client list reconstruction), queries can be issued in an anonymized way (e.g., spatial k-anonymity). The Anonymizer generates a small anonymized spatial region (ASR) for answering queries. Assuming $k = 5$, the ASR is the yellow rectangle that contains five employees (e.g., A, B, C, D, E). Interestingly, the ASR only encompasses employees or ACME; therefore, the anonymized LBS (or curator) is highly sure that the location/route query is issued by the ACME's employees. This example demonstrates that, in some cases, even anonymization methods can violate group privacy. There are many central server-based applications that can either reconstruct the whole data of certain groups or predict sensitive information about a particular group.

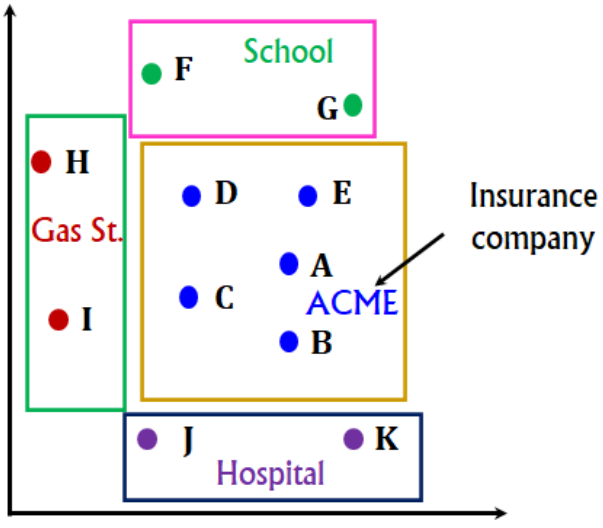


Figure 11. Overview of group privacy breach via movement boundary attack.

Due to the extensive use of SN, groups in the form of communities can easily be detected from SN data based on interest, location, demographic, activities, and profile similarities [139]. Many approaches have been developed to secure the privacy of communities in SN data [140–144]. In recent years, due to the rapid increase in avenues of personal data generation and the availability of analytical tools, group privacy can be compromised easier than individual privacy [145]. Hence, the research question to be addressed regarding group privacy is stated as follows. *How to preserve the privacy of a group/community (i.e., a*

group of people with some common properties/attributes) while guaranteeing the utility of data for analytical and data mining tasks?

The SOTA approaches published in recent years regarding group privacy preservation are summarized as follows. Wu et al. [146] discussed the collective privacy concepts in social networking sites. The authors stressed the need of viewing privacy from the group level rather than the individual level as a common information practice. Reviglio et al. [147] discussed group privacy in the context of big data analytics. The authors described many challenges to group privacy amid technological developments. Gstrein et al. [148] discussed group privacy issues from a technical perspective, and discussed the need for group privacy preservation in the context of algorithmically driven systems. The author discussed the implications of group privacy in decision-making processes. Mavriki et al. [149] discussed group privacy issues that can emanate during the extensive use of big data analytics and mining. The authors described that group privacy issues can have longer implications on groups (i.e., minorities) than individuals, especially in the context of political purposes. Mavriki et al. [150] discussed the group privacy issues in the healthcare sector due to the extensive adoption of mass surveillance and digital technologies (i.e., big data, artificial intelligence, wearables, sensing technologies, etc.). The authors summarized the threats to the individual as well as group privacy in the era of the COVID-19 pandemic, especially in contact tracing applications. Heinrichs et al. [151] discussed the issues of discrimination and hate against groups formed/extracted by artificial intelligence technologies. Authors discussed that AI algorithms can assist in discriminating against people based on their group memberships. Mühlhoff et al. [152] discussed the privacy issues created by predictive data analytics to individual and group privacy, respectively. Authors discussed that group privacy can be easily violated even if individuals provide their data anonymously. Furthermore, statistical inferences and sensitive information prediction leveraging aggregate datasets can cause severe privacy breaches to the groups than individual privacy.

Mavriki et al. [153] discussed the privacy breaches to the groups via profiling using big data. The authors stressed the need of developing group privacy protection methods against big data profiling based on sexuality, health, and race information. Kikuchi et al. [154] discussed the DDP-based solution for the anonymization of transactional data in order to preserve group privacy. The proposed approach strongly preserves the privacy of individuals and transactional data utility. Flood et al. [155] discussed the group privacy issues in the tracing and tracking technologies that were heavily used in the COVID-19 pandemic. The authors stressed the need for the privacy preservation of groups of people who can spread infection due to their activities. Alanezi et al. [156] discussed the group as well as individual privacy preservation in IoT scenarios. The developed prototype can provide strong resilience against both individual and group privacy issues in IoT environments. Wickramasinghe et al. [157] discussed the privacy solution in IoT environments based on individual and group privacy preferences. By using the proposed approach, users in IoT environments can make intelligent strategies regarding their data sharing and collection with privacy guarantees. Kim et al. [158] discussed the emergence of group privacy issues due to the failure of algorithmic transparency. Authors discussed that due to a lack of algorithmic transparency and data processing in a black-box manner, digital environments are prone to group privacy disclosures. Kim et al. [159] discussed the privacy preservation of both groups and individuals in online environments. The authors described that compromising group privacy can often lead to the compromise of an individual's privacy. Hence, protecting group privacy is equally as important as protecting an individual's privacy. Russo et al. [160] developed a blockchain-based privacy-preserving solution for e-commerce applications. The proposed concept works on the principle of data minimization and ensures the privacy protection of users in digital environments. Labs et al. [161] discussed the privacy issues in the context of COVID-19. The authors discussed the strategies of information sharing and corresponding privacy techniques in the COVID-19 era. Alshawi et al. [162] discussed the privacy issues of contact tracing applications employed to control the spread of COVID-19. The authors highlighted the

need for enforcement of special policies in order to control the privacy issues of contact-tracing applications. In addition, the authors highlighted the need to communicating risks to individual privacy at the time of collection. Despite many studies, the research on group privacy topic is still at the early stage, and many technological developments are needed to preserve it in big data environments.

Recently, a comparative analysis of how users perceive privacy and security for group chat was conducted by Sean et al. [163]. The authors surveyed 996 respondents from the UK and the USA with many questions and found out that most users rely on non-technical strategies (i.e., group membership analysis, self-filtering, etc.) in order to preserve their privacy. Petrén et al. [164] discussed the concept of mitigating group disparities through the DP model. The authors concluded that DP could lead to biased results for minority groups in some cases. Erickson et al. [165] provided a privacy analysis of the Femtech app, a technology stack that fulfills the health needs of females. The authors highlighted the privacy breaches in Femtech, and examined the regulatory and technical measures in order to improve the privacy of the app. Perino et al. [166] discussed the vulnerability brought on by AI tools to group privacy. Authors determined that the use of AI tools can introduce a new attack vector in many sectors, especially telco networks. Tadic et al. [167] designed a prototype that preserves the privacy and security of activists online. Sfar et al. [18] proposed generalized privacy-preserving solutions for e-health applications using the game theory concept. Zhang et al. [168] discussed the concept of visual privacy which has become a major threat to the individual as well as group privacy amid rapid developments in AI tools. Wang et al. [169] discussed group privacy issues in next-generation Internet (also known as Metaverse). The authors pinpointed that group privacy issues can be related to a social group, a firm, and even a nation. Nash et al. [170] discussed privacy issues in policy-making concerning an individual or groups. The authors discussed an example of how the collection of data by tech giants such as Facebook violates Australia's privacy principles. The authors also discussed the role of data visualization in the policy-making process. Despite these developments, there is a serious lack of methods that can analyze group privacy issues in the context of both AI and big data. We summarize and compare the famous group privacy protection techniques in Table 1.

Table 1. Summary and comparison of group privacy protection techniques.

Ref.	Study Nature	Main Assertion	Experimental Analysis	Threats to Group Privacy Discussed
Wu et al. [146]	Theoretical	Suggests a method for self-identity protection across social networks	×	×
Reviglio et al. [147]	Theoretical	Highlights pertinent threats to group privacy in data mining	×	✓
Gstrein et al. [148]	Theoretical	Discusses many group privacy issues in datafication paradox	×	✓
Mavriki et al. [149]	Theoretical	Discusses implications of group privacy on general public	×	✓
Mavriki et al. [150]	Technical	Discusses group privacy issues in e-health applications with examples	×	✓
Heinrichs et al. [151]	Theoretical	Highlight group privacy issues caused by the AI tools	×	✓
Mühlhoff et al. [152]	Theoretical	Suggests a group privacy protection in predictive analytics	×	✓
Mavriki et al. [153]	Theoretical	Suggests protecting the interests of groups in big data era	×	✓
Kikuchi et al. [154]	Theoretical	Highlights the need of group privacy protection in purchase records	×	✓
Flood et al. [155]	Theoretical	Discusses group privacy issues in COVID-19 contact tracing apps	×	✓
Alanezi et al. [156]	Technical	Solves the group privacy problem in IoT scenarios using diversity concept	✓	✓
Wickrama et al. [157]	Theoretical	Discusses group privacy issues in smart homes environments	×	✓
Kim et al. [158]	Theoretical	Discusses AI effects on group privacy and their implications	×	✓
Kim et al. [159]	Theoretical	Highlights threats to group privacy in social networks analysis and mining	×	✓
Russo et al. [160]	Technical	Suggests a privacy protection method to access online social network services	✓	×
Labs et al. [161]	Theoretical	Highlights privacy issues in COVID-19 era (surveillance related)	×	✓
Alshawi et al. [162]	Theoretical	Describes group privacy-related issues in COVID-19 tracing apps	×	✓
Sean et al. [163]	Theoretical	Describes privacy concerns in group chat tools	×	✓
Petrén et al. [164]	Technical	Describes DP effect on group privacy (or decisions) preservation	✓	✓
Erickson et al. [165]	Theoretical	Describes many potential group privacy issues in Femtech app	×	✓
Perino et al. [166]	Theoretical	Highlights AI effect on users privacy and change in privacy landscape	×	✓
Tadic et al. [167]	Technical	Develops a practical tool for solving group privacy issues online	✓	✓
Sfar et al. [18]	Technical	Proposed a generalized privacy protection solution for e-health sector	✓	×
Zhang et al. [168]	Theoretical	Highlights the concept of visual privacy in deep learning systems	×	✓
Wang et al. [169]	Theoretical	Provides three new dimensions of group privacy in digitization age	×	✓
Nash et al. [170]	Theoretical	Discusses privacy issues in big data aggregation and analytics	×	✓
This study	Technical	Describes group privacy in AI and big data era and proves concepts' feasibility via experiments	✓	✓

✓⇒ available/reported and ×⇒ not-available/not-reported

As shown in Table 1, most existing studies are theoretical in nature, meaning they provide a generic concept of group privacy. Furthermore, the experimental evaluations (e.g., results) of most existing studies have not been reported in detail. In contrast, our study comprehensively describes the group privacy issues in AI and the big data era that have not been thoroughly discussed in the current literature. Additionally, we demonstrate the utility of our concept with an experimental evaluation that can pave the way for understating this new dimension of privacy in real-world person-specific dataset handling (i.e., aggregation, storage, processing, and distribution).

Threats to the Group Privacy in the Era of AI and Big Data

In this subsection, we discuss the key threats to group privacy in the era of AI and big data. With the introduction of new technologies such as federated learning, swarm learning, big data, and machine and deep learning, the privacy threats landscape has changed from individual to collective. Due to these technologies, the extraction of fine-grained data about the individual as well groups has become relatively easy. Therefore, privacy preservation has become very challenging in recent times. Interestingly, existing research has mainly focused on individual privacy preservation. Therefore, in fact, profiling and AI technologies are targeting group-level analytics and mainly focus on targeting the collective/groups compared to the individual. Although the usage of collected data is beneficial, privacy issues can bring serious consequences, as shown in Figure 12. Hence, the privacy preservation of the group has become equally as important as the individual [171]. Existing research has discussed group privacy issues, but most of them are theoretical in nature. Hence, devising practical solutions for group privacy preservation is a rich area of research.

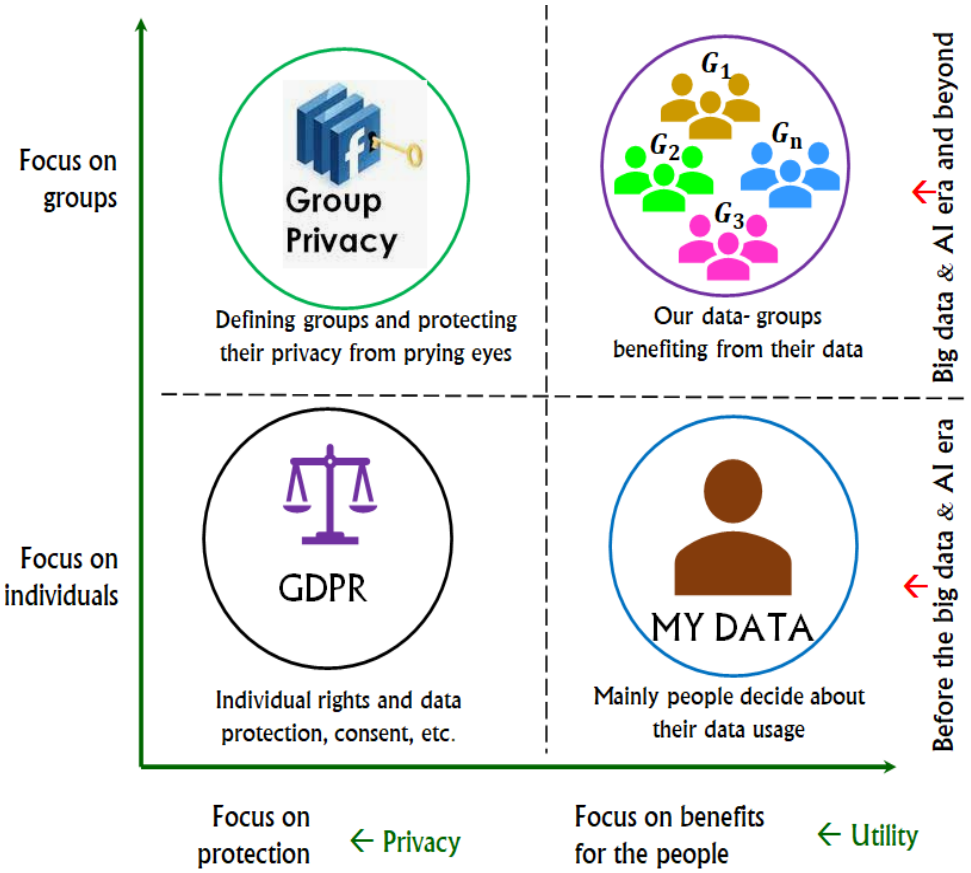


Figure 12. Overview of the changing dynamic of individual and group privacy (bright and dark sides).

Since collective information is a vital source and can be used to discriminate against people or suppress them to meet certain ends, group privacy is therefore an urgent problem to be addressed amid rapid advancements in AI, knowledge-based systems, and big data environments [172]. Recently, there has been a significant lack of group privacy-preserving methods, and group privacy issues are advancing with the passage of time due to the rise in avenues of personal data collection [173,174]. Interestingly, privacy preservation methods used for individual privacy preservation cannot be directly applied to group privacy protection due to the difference in protection goals. For example, *k*-anonymity hides the information of an individual in a group and is a useful solution for individual privacy protection (individual $\rightarrow$ group). However, group privacy cannot be efficiently protected using *k*-anonymity (group $\rightarrow$ individual ($\times$)). Hence, robust methods that can safeguard group privacy are needed in the near future. Research on group privacy is going to become a popular topic in the near future. There have recently been multiple threats to group privacy. These threats can occur either based on the attribute values or by the application of advanced AI techniques that can observe commonalities among differences and vice versa. Based on the extensive analysis of the published literature, we present twenty-five major threats to group privacy as follows.

1. Hidden profiling of group motives/goals;
2. Prediction of norms of particular groups;
3. Inference of political views;
4. Stalking of the groups;
5. Declining credits to the group;
6. Inference of the religious views;
7. Inaccurate and biased decision making about group;
8. Political victimization of minor communities;
9. Spatial-temporal activities disclosures of group;
10. Disclosure of the disease/income of a group;
11. Sensitive rules extraction about minor groups;
12. Collection of intimate details of groups lifestyle;
13. Cyberbullying based on ethnicity of a particular group;
14. Denying fair share in government schemes to a particular group;
15. Community association (or political party association) disclosure;
16. Data aggregation for group privacy theft via statistical matching;
17. Aggregation of social network usage and posted contents' data;
18. Information contagion and control to a particular group;
19. Disclosure of the opinion and sentiments of a group;
20. Harassment of the people due to affiliation with a controversial group;
21. Targeted crime involvement based on presence at some locality via location data;
22. Disclosure of eating or sexual behavior through profiling leveraging common data;
23. Disclosure of common diseases about a group of people living in some parts of the country based on zip code data;
24. Prediction of future motives/activities of a particular group based on historical data;
25. Targeted political surveillance of a certain group.

In addition to the threats cited above, AI-powered attacks can also lead to unexpected privacy breaches for a group of people. Hence, there is an emerging need to devise practical solutions for simultaneously preserving individual and group privacy. In the recent literature, few approaches have been devised to address privacy issues in different computing paradigms. A community privacy preservation leveraging entropy and susceptibility concept was devised by Majeed et al. [175]. Group privacy issues in the ubiquitous computing paradigms were discussed by Politou et al. [176]. A new dimension in privacy (e.g., group privacy $\rightarrow$ collective privacy) was discussed by Mantelero et al. [177]. A practical method to preserve the privacy of web searches in order to hide the group's interest was given by Elovici et al. [178]. Recently, the COVID-19 pandemic accelerated data transition into cyberspace and therefore the scale and scope of privacy breaches concerning an individual

as well as groups are likely to grow in the near future [179,180]. Hence, group privacy is still a new area and requires significant development from both industry and academia in order to preserve the privacy of groups from corporate surveillance technologies.

4. Case Study to Show the Worth of Investigating Group Privacy in Big Data and AI Era Using Real-World Benchmark Datasets

In this section, we present a case study on real-world datasets in order to show the significance of studying group privacy in the big data and AI era. We evaluated the importance of group privacy from two real-world datasets, namely Adults [181] and Diabetes [182], in order to make our analysis valid. Both these datasets are publicly available and have been extensively used in evaluating privacy-preserving mechanisms. The Adults dataset has five quasi-identifiers (QIDs) and one sensitive attribute (SA). Moreover, we removed other non-QID attributes from it. The latter dataset encompasses five QIDs and one SA. The Diabetes 130-US hospitals dataset is a big healthcare dataset related to the diagnosis of diabetes in the US. We used a substantial number of records with five QIDs and one SA from this dataset in our experimental analysis. We present a concise overview of both datasets in Table 2. All datasets were pre-processed (e.g., missing values analysis, outliers removal, format conversion/enrichment, etc.) before actual utilization in the simulation experiments. In the Adults dataset, we performed a *min–max* analysis on age QID by using the values' range information. The *min* and *max* values were 17 and 90, respectively. The *min–max* analysis ensured that all values of age are consistent (e.g., within the desirable range) with the values range (i.e., 17–90) provided by the data owner. Similarly, the *min–max* analysis on the numerical QIDs of diabetes datasets was also performed. After pre-processing, the error-free datasets were used in experimental evaluation.

Table 2. Details of the datasets used in the experiment evaluations.

Dataset	Total Records	Dimensions	QID Name (Cardinality, Type)	Name of SA (Unique Values)
Adults [181]	32,561	32,561 × 6	Age (74, numerical) Gender (2, categorical) Race (5, categorical) Country (41, categorical) Relationship (6, categorical)	Salary/Income (2)
Diabetes [182]	20,501	20,501 × 6	Race (5, categorical) Gender (2, categorical) Age (32, numerical) I_status (4, categorical) Admission_type (8, categorical)	DiabetesMed (2)

Table 3 presents the taxonomy of notations used in the proposed method.

After acquiring the real-world benchmark datasets, D, where D contains N users, we chose a machine learning algorithm named random forest (RF) [183] to identify the group privacy vulnerability based on attribute values. We intend to estimate the privacy risks to Type-*b* (i.e., common features) of group privacy as described by Loi et al. [136], and it was not comprehensively analyzed in the previous literature. In Figure 13, we demonstrate the procedure employed to compute the group privacy risk probability based on attribute values using the RF algorithm. The rationale behind the RF choice for group privacy risks computation is its ability to yield superior accuracy, less difficulty in the specification of the parameters, and proven success in similar tasks (e.g., spam email classification, attributes ranking for credit decisions, protein sequence analysis, COVID-19 tally predictions, search ranking, etc.), and the usage of information-theocratic concepts (i.e., Gini index, Shannon entropy, permutation, etc.). A similar task cannot be done by using decision trees (DTs) due

to lower reliability in their final results (i.e., heavy reliance on just one tree's result that may not be consistent in each domain) and inability to handle the interactions of various kinds between QIDs. Similarly, SVM may not yield consistent results for risk computation when the domain of the SA is higher than the size of the dataset.

Table 3. Taxonomy of notations used in the proposed method.

Symbols	Description
D	Original data
N	Number of individuals in D, where $N = \|D\|$
A	Set of attributes in D, where $A = \{a_1, a_2, \ldots, a_n\}$
u_i	ith user/tuple/record in T, where $t_i = u_i$
Q	Set of QIDs, where $Q = \{QI_1, QI_2, \ldots, QI_p\}$
p	Total QIDs in set Q, where $p = \|Q\|$
S	SA values set, where $S = \{v_1, v_2, \ldots, v_{\|s\|}\}$
$ntree$	Number of classification/regression trees
$mtry$	Variables used to split tree's node
γ_{q_k}	Vulnerability value of a kth QI
γ_D	Total vulnerability of a D
δ	Group of people with some common attributes

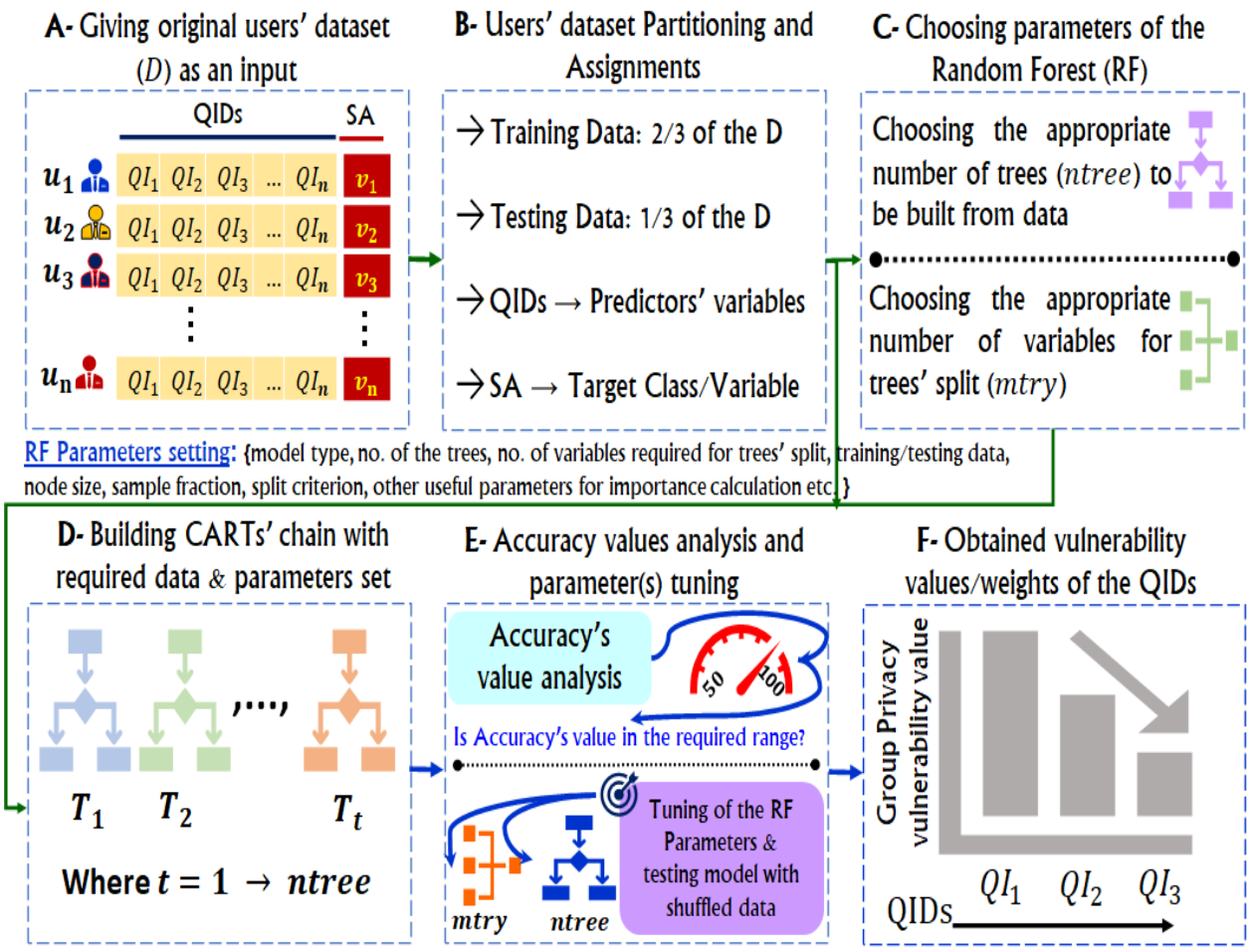


Figure 13. Computing the vulnerability of group privacy based on attribute values using RF.

To compute group privacy risk/vulnerability based on QIDs, six concepts (i.e., $A \rightarrow F$) are applied, as shown in Figure 13. In the first four steps, a chain of classification/regression trees (also known as CARTs) is built from the training data with the parameters specified in the RF formula. The RF model's type can be classification or regression depending upon the SA value. IF the SA is numerical, the RF model type is regression and vice versa. The most critical step in the whole group privacy vulnerability estimation is E in which the accuracy analysis is performed and one reference measure for accuracy is obtained. Later, the values of the QIDs are shuffled (one at a time), and again, the accuracy values are determined by building an RF model again with shuffled QID values together with non-shuffled QIDs. Subsequently, the difference between both accuracies is measured, and attributes risk is computed. If the difference between reference accuracy and newly measured accuracy is not large, it implies that most values of a QID are the same, leading to much higher chances of group privacy disclosures. During the values shuffling process, the QID affect/importance in each tree can be different. Hence, the mean ($\overline{x}_{q_k}$) importance $QIDI$ is calculated for each QID from all trees using Equation (4):

$$\overline{x}_{q_k} = \frac{\sum_{t=1}^{ntree} QIDI^t(q_k)}{ntree} \tag{4}$$

where $\overline{x}_{q_k}$ gives the mean score from all trees. The standard deviation s_{q_k} and vulnerability risk γ_{q_k} can be computed using Equations (5) and (6), respectively.

$$s_{q_k} = \sqrt{\frac{1}{ntree-1}\sum_{t=1}^{ntree}(QIID^t(q_k) - \overline{x}_{q_k})^2} \tag{5}$$

$$\gamma_{q_k} = \frac{\overline{x}_{q_k}}{s_{q_k}} \tag{6}$$

Equation (6) gives the vulnerability value γ for the kth QID present in a dataset.

Based on the γ computing process explained above, attributes can be classified as highly risky, risky, and less risky, respectively. Accordingly, privacy protection can be ensured, taking into account such valuable statistics about attributes from the underlying data.

The RF-based vulnerability computation process for QIDs present in a dataset is the first practical step toward identifying the basis of group privacy breaches using ML. We applied the RF-powered γ computing process on two datasets listed in Table 2, and the corresponding experimental results are shown in Figure 14. From the results, it can be seen that some attributes are more highly vulnerable to group privacy disclosures than others. Interestingly, the top three attributes can expose group privacy up to 66 % in the adult dataset. In contrast, the group privacy vulnerability in the diabetes dataset based on four attributes is approximately 55 %. The higher vulnerability in the adults dataset is mainly due to a higher imbalance and more records compared to the diabetes dataset. This experimental analysis indicates that group privacy is a genuine problem in the context of big data. Furthermore, not only in tabular data but also in SN data, the re-identification of groups is also possible with approximately 80% accuracy in some cases [184]. Hence, rigorous solutions that can offer a solid defense against group privacy problems are needed in the future for responsible data science (https://redasci.org/, Accessed on: 18 March 2022).

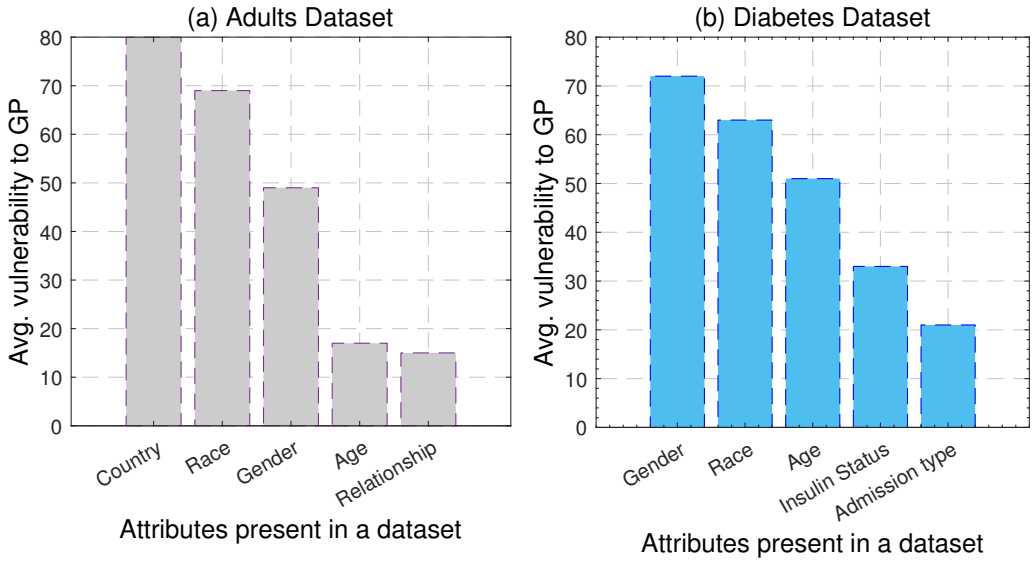


Figure 14. Overview of the vulnerability of group privacy in two datasets based on attributes.

In order to measure the vulnerability in the context of big data, we partitioned the datasets into different chucks. The ten chunks for each dataset and the number of records in each chuck are shown below Equations (7)–(9). For the sake of simplicity, we denote the adults dataset with A, diabetes dataset with D, and hybrid dataset with H. In order to verify the feasibility, we created a hybrid dataset by combining A and D, which is relatively bigger than both A and D. In the hybrid dataset, we kept the distribution of the SA values the same as in the original datasets:

$$A = \{3K, 6K, 9K, 12K, 15K, 18K, 21K, 24K, 27K, 30K, 32.5K\} \tag{7}$$

$$D = \{2K, 4K, 6K, 8K, 10K, 12K, 14K, 16K, 18K, 20K, 20.5K\} \tag{8}$$

$$H = \{5K, 10K, 15K, 20K, 25K, 30K, 35K, 40K, 45K, 50K, 53.5K\} \tag{9}$$

After creating thirty different versions of the datasets, we applied the RF algorithm by choosing appropriate values of it on different versions of data to determine the vulnerability of group privacy. The total vulnerability of group privacy based on all attributes information in each version of the data can be measured using (10).

$$\gamma_D = \sum_{i=1}^{p} \gamma_{q_i} \tag{10}$$

where p denotes the total number of QIDs, and γ_{q_i} denotes the vulnerability of an ith QID.

The experimental analysis obtained from extensive experiments using real-world data and the RF algorithm is shown in Figure 15. As shown in Figure 15, the γ value increases with the number of records. The threats to group privacy can reach up to 80% if a substantial number of records are present in a dataset. These results support our findings of group privacy risks in big data environments using AI/ML techniques. Interestingly, if a dataset is highly imbalanced (e.g., the distribution of values of either sensitive or basic attributes are not uniform), the threats to group privacy can grow in large numbers. As shown in Figure 15 left, due to an imbalance in values of certain attributes, the average vulnerability of group privacy in the adults dataset is relatively higher than the bkseq dataset. These findings and analysis verify our original hypothesis and highlight that group privacy is a genuine issue in big data environments leveraging AI tools.

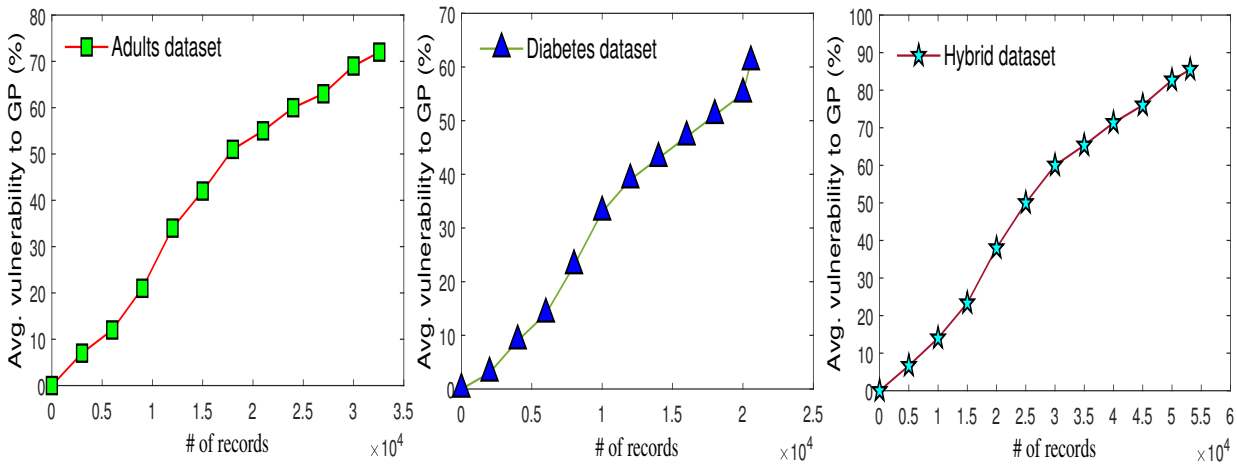
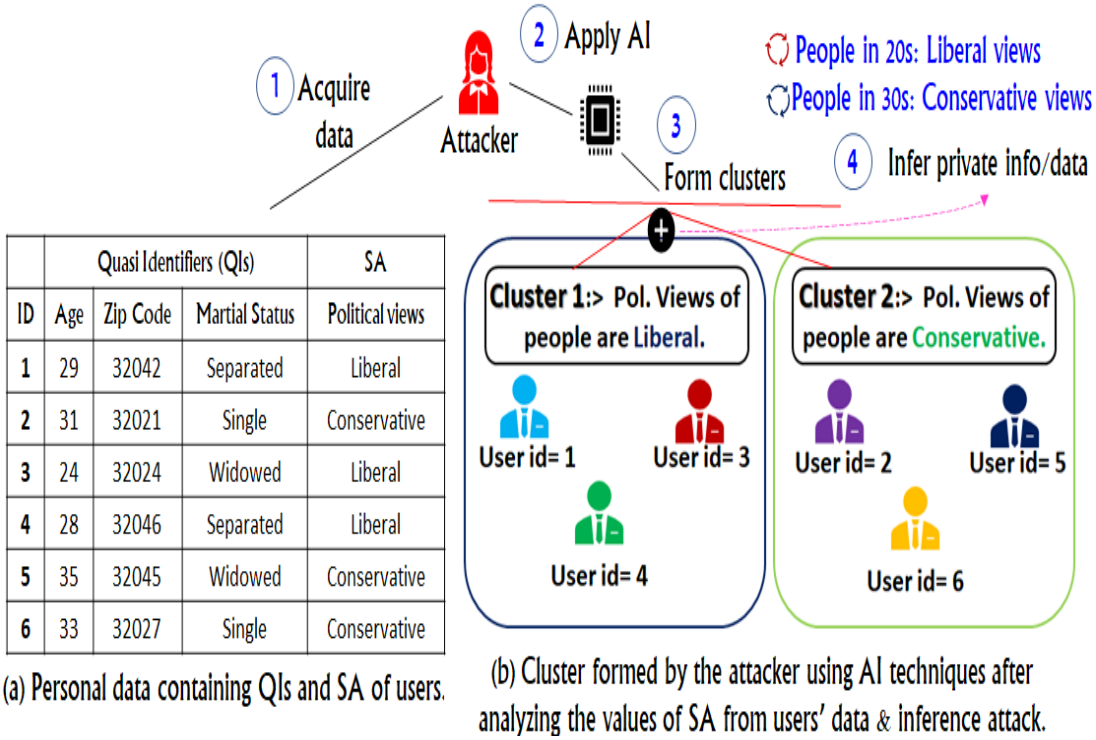


Figure 15. Experimental results for vulnerability of group privacy in real-world datasets.

Apart from the vulnerability risk emerging from the basic information (QIs), in some cases, adversaries can perform clustering based on sensitive information (also known as SA) to compromise group privacy. For example, in an adult dataset, the SA is income. In this SA, only two values (e.g., $\leq$50 K and >50 K) exist, and the frequency of one value is significantly large. In this situation, clustering the records around the dominant value of the SA can expose group privacy. In big data environments, such types of attacks can easily be launched using unsupervised learning techniques (i.e., *k*-means, *k*-medoid, DBSCAN, evolutionary clustering, etc.), and identities or other private information can be inferred. We demonstrate an overview of clustering-based attacks in Figure 16 that can lead to the disclosure of private group information (i.e., political views).



	Quasi Identifiers (QIs)			SA
ID	Age	Zip Code	Martial Status	Political views
1	29	32042	Separated	Liberal
2	31	32021	Single	Conservative
3	24	32024	Widowed	Liberal
4	28	32046	Separated	Liberal
5	35	32045	Widowed	Conservative
6	33	32027	Single	Conservative

(a) Personal data containing QIs and SA of users.

(b) Cluster formed by the attacker using AI techniques after analyzing the values of SA from users' data & inference attack.

Figure 16. Group privacy problems due to SA categories-based clustering attack.

Group Privacy Preservation Results Comparison

The mainstream solution for individual/group privacy preservation in data analysis is anonymization. Since our method identifies the hidden statistics (i.e., γ) of attributes

from D, group privacy issues can be effectively restricted. To compare the performance, we draw a subset of six records from the adults dataset (see Figure 17a). In Figure 17, we compare the results of this study with the chaos and perturbation-based anonymization (CPA) method [185]. The CPA method was proposed to preserve the privacy of big data. We designed three attack scenarios (two commons, and one AI-powered) that can be launched on anonymized data to infer the SA of groups. In the first two scenarios, we assume that an adversary has access to some basic attributes of groups, and they want to infer the SA of the respective group based on association rules. In the third scenario, the adversary trains the AI model with anonymized data and then infers the SA of a particular group in the validation stage. As shown in Figure 17c, this study can increase the degree of generalization when SA lacks heterogeneity to effectively preserve group privacy. In all three representative scenarios, the proposed approach reduces group privacy disclosures by 45.05%. The proposed approach can yield lower group privacy disclosures in any D by making combined use of SA's heterogeneity and QIDs γ information.

In contrast, the CPA method [185] neither identifies the γ information nor considers heterogeneity, and therefore, the probabilistic disclosure at the group level is significantly higher in all three scenarios, as shown in Figure 17b. These results fortify the efficacy of our approach for group privacy preservation against general as well as AI-powered attacks. Our approach can yield expected results (i.e., group privacy preservation) from the whole D as well. Although our approach is a major development, some factors such as data imbalance, noise in data, data-style other than a table, and a substantial number of categories under a single column can degrade its performance in real-world cases.

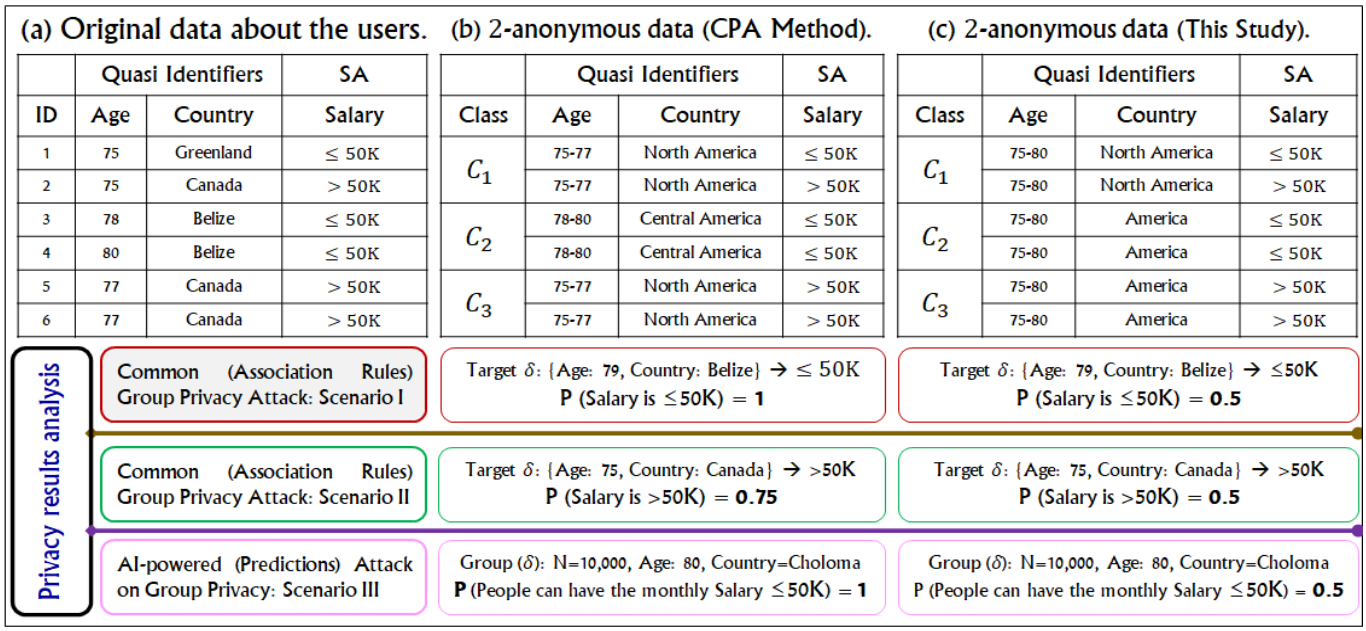


Figure 17. Comparisons of group privacy preservation: this study versus CPA method.

5. Future Research Outlook in the Domain of Privacy Preservation

This section highlights the future research outlook in the domain of privacy preservation from three distinct perspectives: (i) Group privacy preservation in static (i.e., traditional data collection, anonymization, and publishing, published data analytics, etc.) and dynamic (i.e., executing queries and acquiring responses from cloud-based systems) scenarios; (ii) Group privacy preservation in four different computing paradigms; and (iii) Privacy preservation of artificial intelligence (AI) systems/ecosystem.

5.1. Group Privacy Preservation in Static and Dynamic Scenarios

From a group privacy point of view, more technical solutions are required to overcome privacy issues in static as well as dynamic scenarios. In addition, developing hybrid

solutions that can simultaneously ensure individual and group privacy is vital in the era of big data and AI. Furthermore, incorporating AI methods in the design of anonymization to effectively preserve privacy and utility is a vibrant area of research. In addition, exploring group privacy issues and corresponding implications in the era of COVID-19 is one of the promising avenues for future research. Lastly, integrating multidisciplinary approaches with anonymization in order to preserve group privacy is an interesting research area for the coming years. We suggest following three tracks for group privacy in the near future.

- Group privacy protection methods: The concept of group privacy was mainly started in 2014 with theoretical description [186]; until then, many studies were already published to preserve individual privacy. The concept of preserving individual privacy fist emerged in 2002, and there exist many citations of the corresponding studies to date (i.e., <k-anonymity: $8K^+$>, <ℓ-diversity: $6K^+$>, <t-closeness: $3.9K^+$>, < differential privacy:$8.3K^+$>) that proposed ways to protect one person's privacy. In contrast, there have only been a few hundred citations of the studies that have covered group privacy-related concepts. In the future, devising practical methods based on anonymization, DP, blockchain, secure multi-party computation, zero-knowledge proofs, encryption, masking, pseudonymization, etc., is a promising avenue of research. Furthermore, devising practical solutions to preserve group privacy by incorporating the preferences of the group appears to be a likely popular research topic in the near future. Lastly, devising methods that can permit the analysis of data while respecting group privacy is a rich area of research.
- Group privacy evaluation metrics: Since group privacy is a relatively new concept there is therefore a serious lack of metrics that can measure the risks to group privacy in data. In addition, new utility evaluation metrics are also needed to perform the analytics of data while preserving privacy. To this end, metrics that can accurately measure the level of privacy and utility are required in the near future. Furthermore, extending the metrics that were proposed for individual privacy evaluation to group privacy evaluation is also a rich areas of research.
- Group privacy protection in different data styles: Most work in the privacy preservation domain has been done on personal data enclosed in either tabular or graph form. However, personal data can be enclosed in varied forms such as images, text, graphs, videos, streams, matrices, logs, etc. which can lead to group privacy disclosures as well. Hence, devising practical methods to protect group privacy in different data styles is an emerging avenue of research.

5.2. Group Privacy Preservation in Different Computing Paradigms

Apart from the privacy problems in two data styles (e.g., tables and graphs), many other data types such as matrix (e.g., market basket data, trajectories information, transactional databases, and music/movies rating data), digital logs, mobility traces, documents (e.g., medical prescriptions, health recommendations, disease control institutes data), multimedia, text blogs, and temporal data can reveal sensitive information about groups in the digital landscape. Hence, firms and companies are constantly devising new practical strategies and techniques to maintain competitiveness in the market while guaranteeing user/group privacy [187,188]. In the near future, four mainstream technologies such as the Internet of Things (IoT), cloud computing (CC), social networks (SNs), and location-based systems (LBS) will become the core of the information technology (IT) world. These technologies have many benefits such as processing and collecting large-scale datasets in order to identify hidden knowledge [189]. Although these technologies have many benefits, individual and group privacy issues of various kinds can emanate from data processing [190–193]. Hence, these technologies are adopting many privacy-preserving solutions to address this group privacy problem [194–196]. We present potential problems to group privacy in four mainstream technologies (we refer to such technologies as an emerging computing paradigm) after a detailed analysis of the previous studies in Figure 18.

As shown in Figure 18, due to the higher availability of data sources in the emerging computing paradigm, group privacy problems of different kinds can occur. Hence, devising practical privacy-preserving solutions for these paradigms is a promising avenue for future research considering the boom in many of their smart applications. In addition, paying ample attention to the utility aspects of data (also known as drawing pictures out of data) in these technologies also worthily contributes to studying the research problem. Finally, proposing low-cost and flexible privacy-preserving methods that can be applied to more than one data style/format simultaneously in order to preserve both group privacy as well data utility will likely be a rich area of research in future endeavors.

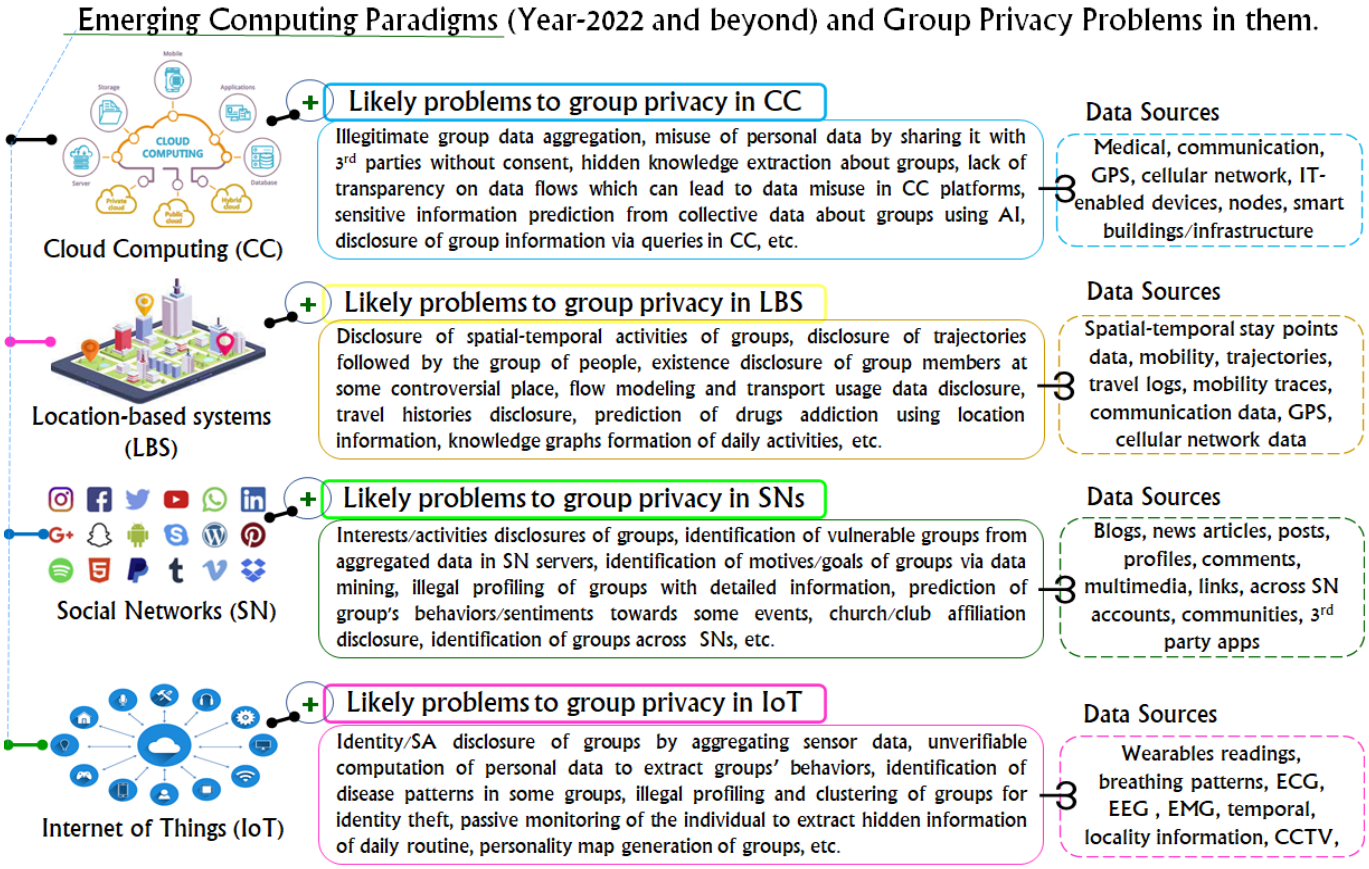


Figure 18. Likely group privacy problems in emerging computing paradigms.

5.3. Privacy Preservation of Artificial Intelligence (AI) Systems/Ecosystem

Recently, the privacy preservation of AI systems, especially federated learning (FL), has become one of the hottest research topics among the research community [197–199]. Due to the distributed nature of FL, FL-based systems are vulnerable to many attacks such as data poisoning, model poisoning, model inversion, gradient inversion, data prediction/reconstruction, and model evasion, to name a few [200]. The key concept of the FL systems is to not centralize data but instead move algorithms close to data. The core difference between centralized learning (CL) and FL is given in the below equation.

$$Case(CL||FL) = \begin{cases} personal\_data \rightarrow algorithms, & \text{CL} \\ algorithms \rightarrow personal\_data, & \text{FL} \end{cases} \tag{11}$$

where CL refers to centralized learning and FL refer to federated learning, respectively.

The privacy of AI systems has been extensively studied across the globe. Despite many developments, this area is relatively new, and challenges need robust solutions [201]. We present an overview of FL, and promising future research directions in Figure 19. The

main research directions listed in Figure 19b are vital to secure AI-based systems from different perspectives.

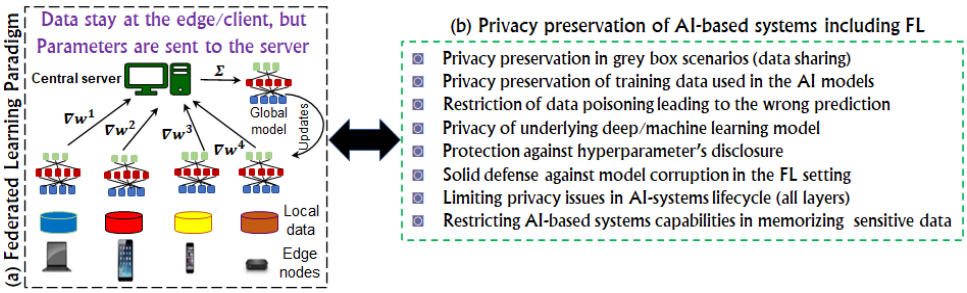


Figure 19. Overview of FL and privacy issues in the context of AI systems.

The main reason to induce the FL concept in this paper is that there are usually multiple clients in the FL ecosystem orchestrated by a central server as shown in Figure 19a. This phenomenon and clients' behavior closely resemble to those of group privacy in traditional data owner settings [202]. In FL environments, the attacker usually takes control of more than one client's data/local model in order to impair the model performance. Similarly, in big data environments, attacks on the group's (e.g., more than one person's) privacy are launched to infer sensitive information of groups. Recently, many techniques have been proposed to secure the clients, server, and aggregation function of the FL ecosystem [203]. Therefore, the techniques that are proposed to secure FL ecosystems such as federated distillation, DP-powered FL for parameters/update security, trusted execution environments (TEEs), zero-knowledge proofs (ZKPs), adversarial training (AT), legitimate participants recognition (LPR), federated multi-task learning, secure multiparty computation (SMC), and confidential computing [204] can be adopted for group privacy preservation in big data and AI environments. Hence, devising new and upgrading existing methods to secure group privacy issues in AI systems is a promising area of research in the near future.

6. Conclusions and Future Work

In this paper, we demonstrated an underrated but highly significant research problem (e.g., group privacy) to be tackled in the recent arena when AI and big data technologies are rapidly advancing. Specifically, we have discussed the major and state-of-the-art developments in individual and group privacy preservation, respectively. We discussed two real-world scenarios to highlight the practicality of the group privacy concept amid the huge proliferation in AI and big data environments in recent times. Furthermore, we experimentally verified the vulnerability of group privacy by applying AI techniques to two real-world benchmark datasets encompassing a substantial number of records. The experimental analysis indicated a rise in the vulnerability of group privacy with an increase in data size as well as data imbalance. These experiments-based findings validate our hypothesis (especially when we talk about groups purposely made by two or more people (with the clear group identifying markers) whose privacy as a group we need to protect), and appear to be well substantiated from both theoretical and practical perspectives. Through experimental analysis, we believe that not only the data size but also some parameter combinations of AI techniques can also lead to fine-grained data derivation/extraction about groups. Adversaries can likely take advantage of the flexibility of hyper-parameters offered by AI techniques to create unexpected privacy breaches. Apart from creating groups and inferring their private information, AI can cause the prediction of sensitive information of particular groups using pre-trained models. Hence, in the near future, AI can expose the hidden characteristics/privacy of groups in big data environments. Therefore, substantial efforts are required from multidisciplinary communities to thwart group privacy problems amid rapid digitization. Finally, we discussed likely threats to group privacy in various emerging computing paradigms (i.e., SNs, CC, IoT, LBS, etc.)

based on the huge proliferation of personal data in them, and suggested potential avenues for future research. The detailed analysis presented in this article can pave the way for developing secure methods in order to preserve group privacy in AI and the big data era. Our work aligns with the recent trends toward responsible data science leveraging AI methods. In the future, we intended to explore federated analytics (FAs) and its significance in the information privacy area. Lastly, we intend to devise a practical anonymization method to safeguard group privacy issues without sacrificing guarantees on data utility in publishing data with researchers/third-party applications.

Author Contributions: All authors contributed equally to this work. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (2020R1A2B5B01002145).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data and studies that were used to support the theoretical and technical findings of this research are included within this article.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Wieringa, J.; Kannan, P.; Ma, X.; Reutterer, T.; Risselada, H.; Skiera, B. Data analytics in a privacy-concerned world. *J. Bus. Res.* **2021**, *122*, 915–925. [CrossRef]
2. Petrescu, M.; Krishen, A. Analyzing the analytics: Data privacy concerns. *J. Mark. Anal.* **2018**, *6*, 41–43. [CrossRef]
3. Vladlena, B. Personal Information Security and the IoT: The Changing Landscape of Data Privacy. *Comput. Commun. Collab.* **2015**, *3*, 15–19.
4. Olakunle, O.; Win, T. Cybersecurity and Data Privacy in the Digital Age: Two Case Examples. In *Handbook of Research on Digital Transformation, Industry Use Cases, and the Impact of Disruptive Technologies*; IGI Global: Pennsylvania, PA, USA, 2022; pp. 117–131.
5. Lu, C.-H.; Chen, Y.-H.; Jan, P.-T. The Privacy Trap of Digital Transformation: The Existence and the Implication. *J. Internet Technol.* **2022**, *23*, 63–71.
6. Florian, V.; Haire, B.; Selvey, L.; Katelaris, A.L.; Kaldor, J. Effectiveness evaluation of digital contact tracing for COVID-19 in New South Wales, Australia. *Lancet Public Health* **2022**, *7*, e250–e258.
7. Hsiang-Yu, Y.; Blakemore, C. The impact of contact tracing and testing on controlling COVID-19 outbreak without lockdown in Hong Kong: An observational study. *Lancet Reg.-Health-West. Pac.* **2022**, *20*, 100374.
8. Jungeun, K.; Rim, S.J.; Jo, M.; Lee, M.G.; Park, S. The Trend of Psychiatric Visits and Psychiatric Medication Prescription Among People Tested for SARS-CoV-2 During the Initial Phase of COVID-19 Pandemic in South Korea. *Psychiatry Investig.* **2022**, *19*, 61.
9. Seoyoung, K.; Lim, H.; Chung, S. How South Korean Internet users experienced the impacts of the COVID-19 pandemic: Discourse on Instagram. *Humanit. Soc. Sci. Commun.* **2022**, *9*, 1–12.
10. Younsik, K. Uncertain future of privacy protection under the Korean public health emergency preparedness governance amid the COVID-19 pandemic. *Cogent Soc. Sci.* **2022**, *8*, 2006393.
11. Kate, E. The Digital Age and Beyond. In *The Routledge Global History of Feminism*; Routledge: London, UK, 2022; pp. 136–148.
12. Jiyoung, S.J.; Metzger, M.J. Privacy Beyond the Individual Level. In *Modern Socio-Technical Perspectives on Privacy*; Springer: Cham, Switzerland, 2022; pp. 91–109.
13. Wadii, B.; Ammar, A.; Benjdira, B.; Koubaa, A. Securing the Classification of COVID-19 in Chest X-ray Images: A Privacy-Preserving Deep Learning Approach. *arXiv* **2022**, arXiv:2203.07728.
14. Miryabelli, A.; Harini, N. Privacy Preservation Using Anonymity in Social Networks. In Proceedings of the Second International Conference on Sustainable Expert Systems, Lalitpur, Nepal, 17–18 September 2021; Springer: Singapore, 2022; pp. 623–631.
15. Naga, P.K.; Rao, M.V.P.C.S. A Comprehensive Assessment of Privacy Preserving Data Mining Techniques. In Proceedings of the Second International Conference on Sustainable Expert Systems, Lalitpur, Nepal, 17–18 September 2021; Springer: Singapore, 2022; pp. 833–842.
16. Yandong, Z.; Lu, R.; Zhang, S.; Guan, Y.; Shao, J.; Wang, F.; Zhu, H. PMRQ: Achieving Efficient and Privacy-Preserving Multi-Dimensional Range Query in eHealthcare. *IEEE Internet Things J.* **2022**. [CrossRef]
17. Oyinlola, O.S. A privacy-preserving multisubset data aggregation scheme with fault resilience for intelligent transportation system. *Inf. Secur. J. A Glob. Perspect.* **2022**, 1–24. [CrossRef]
18. Sfar, A.R.; Natalizio, E.; Mazlout, S.; Challal, Y.; Chtourou, Z. Privacy preservation using game theory in e-health application. *J. Inf. Secur. Appl.* **2022**, *66*, 103158.

19. Quanrun, L.; He, D.; Yang, Z.; Xie, Q.; Choo, K.R. A Lattice-based Conditional Privacy-Preserving Authentication Protocol for the Vehicular Ad Hoc Network. *IEEE Trans. Veh. Technol.* **2022**. [CrossRef]
20. Krishna, P.; Kakade, S.M.; Harchaoui, Z. Robust aggregation for federated learning. *IEEE Trans. Signal Process.* **2022**, *70*, 1142–1154.
21. Balashunmugaraja, B.; Ganeshbabu, T.R. Privacy preservation of cloud data in business application enabled by multi-objective red deer-bird swarm algorithm. *Knowl.-Based Syst.* **2022**, *236*, 107748.
22. Rahul, P.; Patil, P.D.; Kanase, S.; Bhegade, N.; Chavan, V.; Kashetwar, S. System for Analyzing Crime News by Mining Live Data Streams with Preserving Data Privacy. In *Sentimental Analysis and Deep Learning*; Springer: Singapore, 2022; pp. 799–811.
23. Anbar, A.M.A.M.; Manickam, S.; Hasbullah, I.H. A Secure Pseudonym-Based Conditional Privacy-Preservation Authentication Scheme in Vehicular Ad Hoc Networks. *Sensors* **2022**, *22*, 1696.
24. Zhihong, L.; Xing, X.; Qian, J.; Li, H.; Sun, G. Trajectory Privacy Preserving for Continuous LBSs in VANET. *Wirel. Commun. Mob. Comput.* **2022**, *2022*.
25. Huiwen, W.; Wang, L.; Zhang, K.; Li, J.; Luo, Y. A Conditional Privacy-Preserving Certificateless Aggregate Signature Scheme in the Standard Model for VANETs. *IEEE Access* **2022**, *10*, 15605–15618.
26. Nisha, N.; Natgunanathan, I.; Xiang, Y. An Enhanced Location Scattering Based Privacy Protection Scheme. *IEEE Access* **2022**, *10*, 21250–21263. [CrossRef]
27. Kumar, S.; Reddy, S.A.R.; Krishna, B.S.; Rao, J.N.; Kiran, A. Privacy Preserving with Modified Grey Wolf Optimization Over Big Data Using Optimal K Anonymization Approach. *J. Interconnect. Netw.* **2022**, 2141039. [CrossRef]
28. Muhammad, R.S. Hybrid heuristic-based key generation protocol for intelligent privacy preservation in cloud sector. *J. Parallel Distrib. Comput.* **2022**, *163*, 166–180.
29. Kingsleen, S.D.; Kamalakkannan, S. Hybrid optimization-based privacy preservation of database publishing in cloud environment. *Concurr. Comput. Pract. Exp.* **2022**, *34*, e6844. [CrossRef]
30. Wonsuk, K.; Seok, J. Privacy-preserving collaborative machine learning in biomedical applications. In Proceedings of the 2022 International Conference on Artificial Intelligence in Information and Communication (ICAIIC), Riga, Latvia, 21–24 February 2022; IEEE: New York, NY, USA, 2022; pp. 179–183.
31. Joon-Woo, L.; Kang, H.; Lee, Y.; Choi, W.; Eom, J.; Deryabin, M.; Lee, E.; Lee, J.; Yoo, D.; Kim, Y.-S.; et al. Privacy-preserving machine learning with fully homomorphic encryption for deep neural network. *IEEE Access* **2022**, *10*, 30039–30054.
32. Jasmin, Z.; Armknecht, F.; Grohmann, B.; Koch, M. Report: State of the Art Solutions for Privacy Preserving Machine Learning in the Medical Context. *arXiv* **2022**, arXiv:2201.11406.
33. Amin, A.; Shokri, M.; Rabbi, F.; Pun, V.K.I.; Lamo, Y. Extremely Randomized Trees with Privacy Preservation for Distributed Structured Health Data. *IEEE Access* **2022**, *10*, 6010–6027. [CrossRef]
34. Malarvizhi, K.P.; Rawal, B.; Gao, J. Blockchain-enabled Privacy Preserving of IoT Data for Sustainable Smart Cities using Machine Learning. In Proceedings of the 2022 14th International Conference on COMmunication Systems & NETworkS (COMSNETS), Bangalore, India, 4–8 January 2022; IEEE: New York, NY, USA, 2022; pp. 1–6.
35. Arezoo, R.; Ramasubramanian, B.; Maruf, A.A.; Poovendran, R. Privacy-Preserving Reinforcement Learning Beyond Expectation. *arXiv* **2022**, arXiv:2203.10165.
36. Hanchao, K.; Susilo, W.; Zhang, Y.; Liu, W.; Zhang, M. Privacy-Preserving federated learning in medical diagnosis with homomorphic re-Encryption. *Comput. Stand. Interfaces* **2022**, *80*, 103583.
37. Qingyong, W.; Zhou, Y. FedSPL: Federated self-paced learning for privacy-preserving disease diagnosis. *Briefings Bioinform.* **2022**, *23*, bbab498.
38. Zhiyong, H.; Zhou, L.; Zhan, Y.; Liu, C.; Wang, B. Cryptanalysis of an Additively Homomorphic Public Key Encryption Scheme. *Comput. Stand. Interfaces* **2022**, *82*, 103623.
39. Jing, M.; Naas, S.; Sigg, S.; Lyu, X. Privacy-preserving federated learning based on multi-key homomorphic encryption. *Int. J. Intell. Syst.* **2022**. [CrossRef]
40. Sweeney, L. Simple demographics often identify people uniquely. *Health* **2000**, *671*, 1–34.
41. Latanya, S. k-anonymity: A model for protecting privacy. *Int. J. Uncertainty, Fuzziness -Knowl.-Based Syst.* **2002**, *10*, 557–570.
42. Ashwin, M.; Kifer, D.; Gehrke, J.; Venkitasubramaniam, M. l-diversity: Privacy beyond k-anonymity. *Acm Trans. Knowl. Discov. Data (TKDD)* **2007**, *1*, 3-es.
43. Ninghui, L.; Li, T.; Venkatasubramanian, S. t-closeness: Privacy beyond k-anonymity and l-diversity. In Proceedings of the 2007 IEEE 23rd International Conference on data Engineering, Istanbul, Turkey, 17–20 April 2007; IEEE: New York, NY, USA, 2007; pp. 106–115.
44. Yanbing, R.; Li, X.; Miao, Y.; Luo, B.; Weng, J.; Choo, K.R.; Deng, R.H. Towards Privacy-Preserving Spatial Distribution Crowdsensing: A Game Theoretic Approach. *IEEE Trans. Inf. Forensics Secur.* **2022**, *17*, 804–818.
45. Amir, F.; Moghtadaiee, V.; Alishahi, M. On the Privacy Protection of Indoor Location Dataset using Anonymization. *Comput. Secur.* **2022**, *117*, 102665.
46. Joanne, P.; Lineback, J.F.; Bates, N.; Beatty, P. Protecting the Identity of Participants in Qualitative Research. *J. Surv. Stat. Methodol.* **2022**. [CrossRef]
47. Simona, L.E.; Shubina, V.; Niculescu, D. Perturbed-Location Mechanism for Increased User-Location Privacy in Proximity Detection and Digital Contact-Tracing Applications. *Sensors* **2022**, *22*, 687.
48. Abdul, M.; Hwang, S.O. A Practical Anonymization Approach for Imbalanced Datasets. *IT Prof.* **2022**, *24*, 63–69.

49. Ul, I.T.; Ghasemi, R.; Mohammed, N. Privacy-Preserving Federated Learning Model for Healthcare Data. In Proceedings of the 2022 IEEE 12th Annual Computing and Communication Workshop and Conference (CCWC), Las Vegas, NV, USA, 26–29 January 2022; IEEE: New York, NY, USA, 2022; pp. 281–287.
50. Debanjan, S.; Chakraborty, B. Quantifying the Effects of Anonymization Techniques over Micro-databases. *IEEE Trans. Emerg. Top. Comput.* 2022. [CrossRef]
51. Anantaa, K.; Piplai, A.; Chukkapalli, S.S.L.; Joshi, A. PriveTAB: Secure and Privacy-Preserving sharing of Tabular Data. In Proceedings of the ACM International Workshop on Security and Privacy Analytics, Baltimore, MD, USA, 24–27 April 2022.
52. Fengli, X.; Tu, Z.; Li, Y.; Zhang, P.; Fu, X.; Jin, D. Trajectory recovery from ash: User privacy is not preserved in aggregated mobility data. In Proceedings of the 26th International Conference on World Wide Web, Perth, Australia, 3–7 May 2017; pp. 1241–1250.
53. Zhen, T.; Zhao, K.; Xu, F.; Li, Y.; Su, L.; Jin, D. Protecting Trajectory From Semantic Attack Considering *k*-Anonymity, *l*-Diversity, and *t*-Closeness. *IEEE Trans. Netw. Serv. Manag.* **2018**, *16*, 264–278.
54. Soo-Hyun, E.C.; Lee, C.C.; Lee, W.; Leung, C.K. Effective privacy preserving data publishing by vectorization. **Inf. Sci. 2020**, *527*, 311–328.
55. Yang, C.; Xiao, Y.; Xiong, L.; Bai, L. PriSTE: From location privacy to spatiotemporal event privacy. In Proceedings of the 2019 IEEE 35th International Conference on Data Engineering (ICDE), Macao, China, 8–11 April 2019; IEEE: New York, NY, USA, 2019; pp. 1606–1609.
56. Sina, S.; Ding, M.; Liu, B.; Dang, S.; Lin, Z.; Li, J. Privacy preserving location data publishing: A machine learning approach. *IEEE Trans. Knowl. Data Eng.* **2020**, *33*, 3270–3283.
57. José, C.; Pastrana, S. SoK: Privacy-preserving computation techniques for deep learning. *Proc. Priv. Enhancing Technol.* **2021**, *2021*, 139–162.
58. Zhitao, G.; Lv, Z.; Du, X.; Wu, L.; Guizani, M. Achieving data utility-privacy tradeoff in Internet of medical things: A machine learning approach. *Future Gener. Comput. Syst.* **2019**, *98*, 60–68.
59. Farough, A.; Sheikhahmadi, A. DI-Mondrian: Distributed improved Mondrian for satisfaction of the L-diversity privacy model using Apache Spark. *Inf. Sci.* **2021**, *546*, 1–24.
60. Rong, W.; Zhu, Y.; Chen, T.; Chang, C. Privacy-preserving algorithms for multiple sensitive attributes satisfying t-closeness. *J. Comput. Sci. Technol.* **2018**, *33*, 1231–1242.
61. Mehta Brijesh, B.; Rao, U.P. Improved l-diversity: Scalable anonymization approach for privacy preserving big data publishing. *J. King Saud-Univ.-Comput. Inf. Sci.* **2022**, *61*, 1423–1430. [CrossRef]
62. Ullah, B.S.; Jang-Jaccard, J.; Alavizadeh, H. Scalable, high-performance, and generalized subtree data anonymization approach for Apache Spark. *Electronics* **2021**, *10*, 589.
63. Sarah, Z.; Bennani, Y.; Rogovschi, N.; Lyhyaoui, A. Data Anonymization through Collaborative Multi-view Microaggregation. *J. Intell. Syst.* **2021**, *30*, 327–345.
64. Jayapradha, J.; Prakash, M.; Alotaibi, Y.; Khalaf, O.I.; Alghamdi, S. Heap Bucketization Anonymity-An Efficient Privacy-Preserving Data Publishing Model for Multiple Sensitive Attributes. *IEEE Access* **2022**, *10*, 28773–28791. [CrossRef]
65. Satoshi, I.; Kikuchi, H. Estimation of cost of k–anonymity in the number of dummy records. *J. Ambient. Intell. Humaniz. Comput.* **2022**, 1–10. [CrossRef]
66. Cynthia, D. Differential privacy: A survey of results. In Proceedings of the International Conference on Theory and Applications of Models of Computation, Xi'an, China, 25–29 April 2008; Springer: Berlin/Heidelberg, Geramny, 2008; pp. 1–19.
67. Sina, S.; Shamsabadi, A.S.; Bellet, A.; Gatica-Perez, D. GAP: Differentially Private Graph Neural Networks with Aggregation Perturbation. *arXiv* **2022**, arXiv:2203.00949.
68. Jiawen, Q.; Wang, J.; Li, Q.; Fang, S.; Li, X.; Lei, L. Differentially private frequent episode mining over event streams. *Eng. Appl. Artif. Intell.* **2022**, *110*, 104681.
69. Ding, W.; Yang, W.; Zhou, J.; Shi, L.; Chen, G. Privacy Preserving via Secure Summation in Distributed Kalman Filtering. *IEEE Trans. Control. Netw. Syst.* **2022**. [CrossRef]
70. Ahmed, E.O.; Abdelhadi, A. Differential Privacy for Deep and Federated Learning: A Survey. *IEEE Access* **2022**, *10*, 22359–22380.
71. Yanbing, R.; Li, X.; Miao, Y.; Deng, R.; Weng, J.; Ma, S.; Ma, J. DistPreserv: Maintaining User Distribution for Privacy-Preserving Location-Based Services. *IEEE Trans. Mob. Comput.* **2022**. [CrossRef]
72. Seira, H.; Murakami, T. Degree-Preserving Randomized Response for Graph Neural Networks under Local Differential Privacy. *arXiv* **2022**, arXiv:2202.10209.
73. Cai, Z.; He, Z. Trading private range counting over big IoT data. In Proceedings of the 2019 IEEE 39th International Conference on Distributed Computing Systems (ICDCS), Dallas, TX, USA, 7–10 July 2019; IEEE: New York, NY, USA, 2019; pp. 144–153.
74. Xu, Z.; Cai, Z. Privacy-preserved data sharing towards multiple parties in industrial IoTs. *IEEE J. Sel. Areas Commun.* **2020**, *38*, 968–979.
75. Yan, H.; Meng, C.; Li, R.; Jing, T. An overview of privacy preserving schemes for industrial internet of things. *China Commun.* **2020**, *17*, 1–18.
76. Eugene, B.; Poursaeed, O.; Shmatikov, V. Differential privacy has disparate impact on model accuracy. In Proceedings of the Advances in Neural Information Processing Systems 32: Annual Conference on Neural Information Processing Systems 2019, NeurIPS 2019, Vancouver, BC, Canada, 8–14 December 2019.

77. Wang, T.; Li, N.; Somesh, J. Locally differentially private frequent itemset mining. In Proceedings of the 2018 IEEE Symposium on Security and Privacy (SP), Francisco, CA, USA, 20–24 May 2018; IEEE: New York, NY, USA, 2018; pp. 127–143.
78. Tao, L.; Lin, L. Anonymousnet: Natural face de-identification with measurable privacy. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, Long Beach, CA, USA, 16–17 June 2019.
79. Celestine, I.; Moqurrab, S.A.; Anjum, A.; Khan, S.; Mohan, S.; Srivastava, G. N-Sanitization: A semantic privacy-preserving framework for unstructured medical datasets. *Comput. Commun.* **2020**, *161*, 160–171.
80. Andreas, N.; Jiménez, A.; Treiber, A.; Kolberg, J.; Jasserand, C.; Kindt, E.; Delgado, H.; Todisco, M.; Hmani, M.A.; Mtibaa, A.; et al. Preserving privacy in speaker and speech characterisation. *Comput. Speech Lang.* **2019**, *58*, 441–480.
81. Sagar, S.; Chen, K.; Sheth, A. Toward practical privacy-preserving analytics for IoT and cloud-based healthcare systems. *IEEE Internet Comput.* **2018**, *22*, 42–51.
82. Ye, Q.; Hu, H.; Meng, X.; Zheng, H. PrivKV: Key-value data collection with local differential privacy. In Proceedings of the 2019 IEEE Symposium on Security and Privacy (SP), Francisco, CA, USA, 19–23 May 2019; IEEE: New York, NY, USA, 2019; pp. 317–331.
83. Hangyu, Z.; Zhang, H.; Jin, Y. From federated learning to federated neural architecture search: A survey. *Complex Intell. Syst.* **2021**, *7*, 639–657.
84. Meng, H.; Li, H.; Xu, G.; Liu, S.; Yang, H. Towards efficient and privacy-preserving federated deep learning. In Proceedings of the ICC 2019–2019 IEEE International Conference on Communications (ICC), Shanghai, China, 20–24 May 2019; IEEE: New York, NY, USA, 2019; pp. 1–6.
85. Hangyu, Z.; Xu, J.; Liu, S.; Jin, Y. Federated learning on non-IID data: A survey. *Neurocomputing* **2021**, *465*, 371–390.
86. Hao, J.; Luo, Y.; Li, P.; Mathew, J. A review of secure and privacy-preserving medical data sharing. *IEEE Access* **2019**, *7*, 61656–61669.
87. Mohamed, S.; Tandon, R.; Li, M. Wireless federated learning with local differential privacy. In Proceedings of the 2020 IEEE International Symposium on Information Theory (ISIT), Los Angeles, CA, USA, 21–26 June 2020; IEEE: New York, NY, USA, 2020; pp. 2604–2609.
88. Odeyomi, O.T. Differential Privacy in Social Networks Using Multi-Armed Bandit. *IEEE Access* **2022**, *10*, 11817–11820. [CrossRef]
89. Tian, W.; Mei, Y.; Jia, W.; Zheng, X.; Wang, G.; Xie, M. Edge-based differential privacy computing for sensor-cloud systems. *J. Parallel Distrib. Comput.* **2020**, *136*, 75–85.
90. Mengnan, B.; Wang, Y.; Cai, Z.; Tong, X. A privacy-preserving mechanism based on local differential privacy in edge computing. *Chin. Commun.* **2020**, *17*, 50–65.
91. Mengmeng, Y.; Zhu, T.; Liu, B.; Xiang, Y.; Zhou, W. Machine learning differential privacy with multifunctional aggregation in a fog computing architecture. *IEEE Access* **2018**, *6*, 17119–17129.
92. Akbar, H.M.; Anwar, A.; Chakrabortty, R.K.; Doss, R.; Ryan, M.J. Differential Privacy for IoT-Enabled Critical Infrastructure: A Comprehensive Survey. *IEEE Access* **2021**, *9*, 153276–153304.
93. Bin, J.; Li, J.; Yue, G.; Song, H. Differential Privacy for Industrial Internet of Things: Opportunities, Applications, and Challenges. *IEEE Internet Things J.* **2021**, *8*, 10430–10451.
94. Mahawaga, A.P.C.; Bertok, P.; Khalil, I.; Liu, D.; Camtepe, S.; Atiquzzaman, M. Local differential privacy for deep learning. *IEEE Internet Things J.* **2019**, *7*, 5827–5842. [CrossRef]
95. Pathum, C.M.A.; Bertok, P.; Khalil, I.; Liu, D.; Camtepe, S. Privacy preserving distributed machine learning with federated learning. *Comput. Commun.* **2021**, *171*, 112–125.
96. Will, A.; Hall, A.J.; Papadopoulos, P.; Pitropakis, N.; Buchanan, W.J. A distributed trust framework for privacy-preserving machine learning. In Proceedings of the International Conference on Trust and Privacy in Digital Business, Bratislava, Slovakia, 14–17 September 2020; Springer: Cham, Switzerland, 2020; pp. 205–220.
97. Chandra, T.; Chamikara, M.A.P.; Camtepe, S.A. Advancements of federated learning towards privacy preservation: From federated learning to split learning. In *Federated Learning Systems*; Springer: Cham, Switzerland, 2021; pp. 79–109.
98. Teng, W.; Zhao, J.; Hu, Z.; Yang, X.; Ren, X.; Lam, K. Local Differential Privacy for data collection and analysis. *Neurocomputing* **2021**, 426, 114–133.
99. Ge, Y.; Wang, S.; Wang, H. Federated learning with personalized local differential privacy. In Proceedings of the 2021 IEEE 6th International Conference on Computer and Communication Systems (ICCCS), Chengdu, China, 23–26 April 2021; IEEE: New York, NY, USA, 2021; pp. 484–489.
100. Veronika, S.; Chamikara, M.A.P.; Khalil, I.; Atiquzzaman, M. Privacy-preserving location data stream clustering on mobile edge computing and cloud. *Inf. Syst.* **2021**, 101728.
101. Milan, L.; Alishahi, M.; Kivits, J.; Klarenbeek, J.; Velde, G.; Zannone, N. Comparing classifiers' performance under differential privacy. In Proceedings of the International Conference on Security and Cryptography (SECRYPT), Lisbon, Portugal, 6–8 July 2021.
102. Seryne, R.; Laurent, M.; Masmoudi, S.; Roux, C.; Mazeau, B. A Validated Privacy-Utility Preserving Recommendation System with Local Differential Privacy. *arXiv* **2021**, arXiv:2109.11340.
103. Afsoon, A.; Mohammadi, B. A clustering-based anonymization approach for privacy-preserving in the healthcare cloud. *Concurr. Comput. Pract. Exp.* **2022**, *34*, e6487.

104. Yavuz, C.; Sagiroglu, S.; Vural, Y. A new utility-aware anonymization model for privacy preserving data publishing. *Concurr. Comput. Pract. Exp.* **2022**, *34*, e6808.
105. Farough, A.; Khamforoosh, K.; Sheikhahmadi, A.; Khamfroush, H. DHkmeans-*l*-diversity: Distributed hierarchical *K*-means for satisfaction of the *l*-diversity privacy model using Apache Spark. *J. Supercomput.* **2022**, *78*, 2616–2650.
106. Rabeeha, F.; Shah, M.A.; Khattak, H.A.; Rauf, H.T.; Al-Turjman, F. Achieving data privacy for decision support systems in times of massive data sharing. *Clust. Comput.***2022**, 1–13. [CrossRef]
107. Ullah, S.F.; Yahya, A. *Clustering Techniques for Image Segmentation*; Springer: Berlin/Heidelberg, Germany, 2022.
108. Kun, G.; Zhang, Q. Fast clustering-based anonymization approaches with time constraints for data streams. *Knowl.-Based Syst.* **2013**, *46*, 95–108.
109. Andrew, O.J.; Karthikeyan, J.; Sei, Y. An efficient clustering-based anonymization scheme for privacy-preserving data collection in IoT based healthcare services. *Peer-to-Peer Netw. Appl.* **2021**, *14*, 1629–1649.
110. Ugur, S.; Abul, O. Classification utility aware data stream anonymization. *Appl. Soft Comput.* **2021**, *110*, 107743.
111. Lu, Y.; Chen, X.; Luo, Y.; Lan, X.; Wang, W. IDEA: A utility-enhanced approach to incomplete data stream anonymization. *Tsinghua Sci. Technol.* **2021**, *27*, 127–140.
112. Sadeghi, N.A.R.; Ghaffarian, H. A New Fast Framework for Anonymizing IoT Stream Data. In Proceedings of the 2021 5th International Conference on Internet of Things and Applications (IoT), Isfahan, Iran, 19–20 May 2021; IEEE: New York, NY, USA, 2021; pp. 1–5.
113. Jimmy, T.; Bouna, B.A.; Issa, Y.B.; Kamradt, M.; Haraty, R. (k, l)-Clustering for Transactional Data Streams Anonymization. In Proceedings of the International Conference on Information Security Practice and Experience, Tokyo, Japan, 25–27 September 2018; Springer: Cham, Switzerland, 2018; pp. 544–556.
114. Pooja, P. Clustering Approaches for Anonymizing High-Dimensional Sequential Activity Data. Ph.D. Thesis, University of Maryland, Baltimore County, MD, USA, 2020.
115. Naixuan, G.; Yang, M.; Gong, Q.; Chen, Z.; Luo, J. Data anonymization based on natural equivalent class. In Proceedings of the 2019 IEEE 23rd International Conference on Computer Supported Cooperative Work in Design (CSCWD), Porto, Portugal, 6–8 May 2019; IEEE: New York, NY, USA, 2019; pp. 22–27.
116. Wantong, Z.; Wang, Z.; Lv, T.; Ma, Y.; Jia, C. K-anonymity algorithm based on improved clustering. In Proceedings of the International Conference on Algorithms and Architectures for Parallel Processing,Guangzhou, China, 15–17 November 2018; Springer: Cham, Switzerland, 2018; pp. 462–476.
117. Madhuri, S.; Li, Y.; Cheng, X.; Tian, Z.; Cai, Z. Anonymization in online social networks based on enhanced equi-cardinal clustering. *IEEE Trans. Comput. Soc. Syst.* **2019**, *6*, 809–820.
118. Zhao, X.; Pi, D.; Chen, J. Novel trajectory privacy-preserving method based on clustering using differential privacy. *Expert Syst. Appl.* **2020**, *149*, 113241. [CrossRef]
119. Qi, L.; Yu, J.; Han, J.; Yao, X. Differentially private and utility-aware publication of trajectory data. *Expert Syst. Appl.* **2021**, *180*, 115120.
120. Yan, X.; Zhou, Y.; Huang, F.; Wang, X.; Yuan, P. Privacy protection method of power metering data in clustering based on differential privacy. In Proceedings of the 2021 IEEE 4th International Electrical and Energy Conference (CIEEC), Wuhan, China, 28–30 May 2021; IEEE: New York, NY, USA, 2021; pp. 1–6.
121. Lan, Q.; Ma, J.; Yan, Z.; Li, G. Utility-preserving differentially private skyline query. *Expert Syst. Appl.* **2022**, *187*, 115871. [CrossRef]
122. Jiawen, D.; Pi, Y. Research on Privacy Protection Technology of Mobile Social Network Based on Data Mining under Big Data. *Secur. Commun. Netw.* **2022**, *2022*, 3826126.
123. Vartika, P.; Kaur, P.; Sachdeva, S. Efficient Clustering of Transactional Data for Privacy-Preserving Data Publishing. In *Cyber Security and Digital Forensics*; Springer: Singapore, 2022; pp. 153–160.
124. Oleksii, P.; Mushkatblat, V.; Kaplan, A. Privacy Attacks Based on Correlation of Dataset Identifiers: Assessing the Risk. In Proceedings of the 2022 IEEE 12th Annual Computing and Communication Workshop and Conference (CCWC), Las Vegas, NV, USA, 26–29 January 2022; IEEE: New York, NY, USA, 2022; pp. 808–815.
125. Ariel, F. Anonymizing Machine Learning Models. In Proceedings of the Data Privacy Management, Cryptocurrencies and Blockchain Technology: ESORICS 2021 International Workshops, DPM 2021 and CBT 2021, Darmstadt, Germany, 8 October 2021; Revised Selected Papers; Springer Nature: Berlin, Germany, 2021, p. 121.
126. Mina, S.; Saracino, A.; Martinelli, F.; Marra, A.L. Privacy preserving data sharing and analysis for edge-based architectures. *Int. J. Inf. Secur.* **2022**, *21*, 79–101.
127. Riyazuddin, M.D.; Begum, S.H.; Sadiq, J. Preserving the Privacy of COVID-19 Infected Patients Data Using a Divergent-Scale Supervised Learning for Publishing the Informative Data. In *Contactless Healthcare Facilitation and Commodity Delivery Management During COVID-19 Pandemic*; Springer: Singapore, 2022; pp. 35–47.
128. Shree, Y.U.; Gupta, B.B.; Peraković, D.; Peñalvo, F.J.G.; Cvitić, I. Security and Privacy of Cloud-Based Online Online Social Media: A Survey. In *Sustainable Management of Manufacturing Systems in Industry 4.0*; Springer: Cham, Switzerland, 2022; pp. 213–236.
129. Tânia, C.; Moniz, N.; Faria, P.; Antunes, L. Survey on Privacy-Preserving Techniques for Data Publishing. *arXiv* **2022**, arXiv:2201.08120.

130. Dong, L.; Yang, G.; Wang, Y.; Jin, H.; Chen, E. How to Protect Ourselves from Overlapping Community Detection in Social Networks. *IEEE Trans. Big Data* **2022**.
131. Chenguang, W.; Tianqing, Z.; Xiong, P.; Ren, W.; Choo, K.R. A privacy preservation method for multiple-source unstructured data in online social networks. *Comput. Secur.* **2022**, *113*, 102574.
132. Srivatsan, S.; Maheswari, N. Privacy Preservation in Social Network Data using Evolutionary Model. *Mater. Today Proc.* **2022**, *in press*. [CrossRef]
133. Shakir, K.; Saravanan, V.; Lakshmi, T.J.; Deb, N.; Othman, N.A. Privacy Protection of Healthcare Data over Social Networks Using Machine Learning Algorithms. *Comput. Intell. Neurosci.* **2022**, *2022*, 9985933.
134. Linnet, T.; Floridi, L.; der Sloot, B.V. (Eds.) *Group Privacy: New Challenges of Data Technologies*; Springer: Berlin/Heidelberg, Germany, 2016; Volume 126.
135. Lanah, K.; Baar, T.; Berens, J.; Letouzé, E.; Manske, J.; Palmer, J.; Sangokoya, D.; Vinck, P. Group privacy in the age of big data. In *Group Privacy*; Springer: Cham, Switzerland, 2017; pp. 37–66.
136. Michele, L.; Christen, M. Two concepts of group privacy. *Philos. Technol.* **2020**, *33*, 207–224.
137. Nora, M.; Forte, A. Privacy and Vulnerable Populations. In *Modern Socio-Technical Perspectives on Privacy*; Springer: Cham, Switzerland, 2022; pp. 337–363.
138. Shabnam, N.; Delic, A.; Tkalcic, M.; Tintarev, N. Factors influencing privacy concern for explanations of group recommendation. In Proceedings of the 29th ACM Conference on User Modeling, Adaptation and Personalization, Utrecht, The Netherlands, 21–25 June 2021; pp. 14–23.
139. Lili, N.Z.; Hrgarek, L.; Welzer, T.; Hölbl, M. Models of Privacy and Disclosure on Social Networking Sites: A Systematic Literature Review. *Mathematics* **2022**, *10*, 146.
140. Xu, Z.; Cai, Z.; Luo, G.; Tian, L.; Bai, X. Privacy-preserved community discovery in online social networks. *Future Gener. Comput. Syst.* **2019**, *93*, 1002–1009.
141. Jin, B.; Li, S. Research on a privacy preserving clustering method for social network. In Proceedings of the 2019 IEEE 4th International Conference on Cloud Computing and Big Data Analysis (ICCCBDA), Chengdu, China, 12–15 April 2019; IEEE: New York, NY, USA, 2019; pp. 29–33.
142. Guobin, C.; Huang, T. Community privacy estimation method based on key node method in space social Internet of Things. *Int. J. Distrib. Sens. Netw.* **2019**, *15*, 1550147719883131.
143. Jian, L.; Zhang, X.; Liu, J.; Gao, L.; Zhang, H.; Feng, Y. Large-Scale Social Network Privacy Protection Method for Protecting K-Core. *Int. J. Netw. Secur.* **2021**, *23*, 612–622.
144. Zengyang, S.; Ma, L.; Lin, Q.; Li, J.; Gong, M.; Nandi, A.K. PMCDM: Privacy-preserving multiresolution community detection in multiplex networks. *Knowl.-Based Syst.* **2022**, *244*, 108542.
145. de, M.Y.; Quoidbach, J.; Robic, F.; Pentland, A.S. Predicting personality using novel mobile phone-based metrics. In Proceedings of the International Conference on Social Computing, Behavioral-Cultural Modeling, and Prediction, Washington, DC, USA, 2–5 April 2013; Springer: Berlin/Heidelberg, Germany, 2013; pp. 48–55.
146. Wu, P.F. The privacy paradox in the context of online social networking: A self-identity perspective. *J. Assoc. Inf. Sci. Technol.* **2019**, *70*, 207–217. [CrossRef]
147. Urbano, R.; Alunge, R. I am datafied because we are datafied": An Ubuntu perspective on (relational) privacy. *Philos. Technol.* **2020**, *33*, 595–612.
148. Gstrein, O.J.; Beaulieu, A. How to protect privacy in a datafied society? A presentation of multiple legal and conceptual approaches. *Philos. Technol.* **2022**, *35*, 1–38. [CrossRef] [PubMed]
149. Paola, M.; Karyda, M. Big data in political communication: Implications for group privacy. *Int. J. Electron. Gov.* **2019**, *11*, 289–309.
150. Paola, M.; Karyda, M. Big Data Analytics in Healthcare Applications: Privacy Implications for Individuals and Groups and Mitigation Strategies. In *European, Mediterranean, and Middle Eastern Conference on Information Systems*; Springer: Cham, Switzerland, 2020; pp. 526–540.
151. Heinrichs, B. Discrimination in the age of artificial intelligence. *AI Soc.* **2022**, *37*, 143–154. [CrossRef]
152. Rainer, M. Predictive privacy: Towards an applied ethics of data analytics. *Ethics Inf. Technol.* **2021**, *23*, 675–690.
153. Paola, M.; Karyda, M. Profiling with Big Data: Identifying Privacy Implication for Individuals, Groups and Society. In Proceedings of the MCIS 2018 Proceedings, Corfu, Greece, 28–30 September 2018; p. 4.
154. Hiroaki, K. Differentially private profiling of anonymized customer purchase records. In *Data Privacy Management, Cryptocurrencies and Blockchain Technology*; Springer: Cham, Switzerland, 2020; pp. 19–34.
155. John, F.; Lewis, M. Monitoring the R-Citizen in the Time of COVID-19. In *Communicating COVID-19*; Palgrave Macmillan: Cham, Switzerland, 2021; pp. 345–370.
156. Khaled, A.; Mishra, S. Incorporating individual and group privacy preferences in the internet of things. *J. Ambient. Intell. Humaniz. Comput.* **2021**, *13*, 1969–1984
157. Ishara, W.C.; Reinhardt, D. A User-Centric Privacy-Preserving Approach to Control Data Collection, Storage, and Disclosure in Own Smart Home Environments. In Proceedings of the International Conference on Mobile and Ubiquitous Systems: Computing, Networking, and Services, Virtual Event, 8–11 November 2021; Springer: Cham, Switzerland, 2021; pp. 190–206.
158. Kitae, K.; Moon, S. When Algorithmic Transparency Failed: Controversies Over Algorithm-Driven Content Curation in the South Korean Digital Environment. *Am. Behav. Sci.* **2021**, *65*, 847–862.

159. Jongwoo, K.; Baskerville, R.L.; Ding, Y. Breaking the privacy kill chain: Protecting individual and group privacy online. *Inf. Syst. Front.* **2020**, *22*, 171–185.
160. Antonia, R.; Lax, G.; Dromard, B.; Mezred, M. A System to Access Online Services with Minimal Personal Information Disclosure. *Inf. Syst. Front.* **2021**, 1–13.
161. Jennifer, L.; Terry, S. Privacy in the coronavirus era. *Genet. Test. Mol. Biomarkers* **2020**, *24*, 535–536.
162. Amany, A.; Al-Razgan, M.; AlKallas, F.H.; Suhaim, R.A.B.; Al-Tamimi, R.; Alharbi, N.; AlSaif, S.O. Data privacy during pandemics: A systematic literature review of COVID-19 smartphone applications. *PeerJ Comput. Sci.* **2022**, *7*, e826.
163. Sean, O.; Abu-Salma, R.; Diallo, O.; Krämer, J.; Simmons, J.; Wu, J.; Ruoti, S. User Perceptions of Security and Privacy for Group Chat. *Digit. Threat. Res. Pract. (Dtrap)* **2022**, *3*, 1–29.
164. Victor, P.B.H.; Neerkaje, A.T.; Sawhney, R.; Flek, L.; Søgaard, A. The Impact of Differential Privacy on Group Disparity Mitigation. *arXiv* **2022**, arXiv:2203.02745
165. Jacob, E.; Yuzon, J.Y.; Bonaci, T. What You Don't Expect When You're Expecting: Privacy Analysis of Femtech. *IEEE Trans. Technol. Soc.* **2022**. [CrossRef]
166. Diego, P.; Katevas, K.; Lutu, A.; Marin, E.; Kourtellis, N. "Privacy-preserving AI for future networks. *Commun. ACM* **2022**, *65*, 52–53.
167. Borislav, T.; Rohde, M.; Randall, D.; Wulf, V. Design Evolution of a Tool for Privacy and Security Protection for Activists Online: Cyberactivist. *Int. J. Hum. Comput. Interact.* **2022**, 1–23.
168. Zhang, G.; Liu, B.; Zhu, T.; Zhou, A.; Zhou, W. Visual privacy attacks and defenses in deep learning: A survey. *Artif. Intell. Rev.* **2022**, 1–55. [CrossRef]
169. Wang, Y.; Su, Z.; Zhang, N.; Liu, D.; Xing, R.; Luan, T.H.; Shen, X. A Survey on Metaverse: Fundamentals, Security, and Privacy. *arXiv* **2022**, arXiv:2203.02662.
170. Kathryn, N.; Trott, V.; Allen, W. The politics of data visualisation and policy making. *Convergence* **2022**, *28*, 3–12.
171. Linnet, T.; Floridi, L.; van der Sloot, B. Introduction: A new perspective on privacy. In *Group Privacy*; Springer: Cham, Switzerland, 2017; pp. 1–12.
172. Hitoshi, K. When accurate information harms people: Information on COVID-19 infection clusters in Japan. *Cosmop. Civ. Soc. Interdiscip. J.* **2021**, *13*, 60–72.
173. Paola, M.; Karyda, M. Automated data-driven profiling: Threats for group privacy. *Inf. Comput. Secur.* 2019. [CrossRef]
174. Murali, K.S.; Kumar, A.P.S. Modern Privacy Threats and Privacy Preservation Techniques in Data Analytics. In *Factoring Ethics in Technology, Policy Making and Regulation*; IntechOpen: London, UK, 2021.
175. Abdul, M.; Lee, S. Attribute susceptibility and entropy based data anonymization to improve users community privacy and utility in publishing data. *Appl. Intell.* **2020**, *50*, 2555–2574.
176. Eugenia, P.; Alepis, E.; Virvou, M.; Patsakis, C. Privacy in Ubiquitous Mobile Computing. In *Privacy and Data Protection Challenges in the Distributed Era*; Springer: Cham, Switzerland, 2022; pp. 93–131.
177. Mantelero, A. From group privacy to collective privacy: Towards a new dimension of privacy and data protection in the big data era. In *Group Privacy*; Springer: Cham, Switzerland, 2017; pp. 139–158.
178. Yuval, E.; Shapira, B.; Maschiach, A. A new privacy model for hiding group interests while accessing the web. In Proceedings of the 2002 ACM workshop on Privacy in the Electronic Society, Washington, DC, USA, 21 November 2002; pp. 63–70.
179. Samuel, R.; Saura, J.R.; Palacios-Marqués, D. Towards a new era of mass data collection: Assessing pandemic surveillance technologies to preserve user privacy. *Technol. Forecast. Soc. Chang.* **2021**, 167, 120681.
180. Paula, G.; López, A.F.; Lacárcel, F.J.S. Main Government-Related Data Extraction Techniques: A Review. *Handb. Res. Artif. Intell. Gov. Pract. Process.* **2022**, 142–160.
181. Blake, C.; Merz, C. *UCI Repository of Machine Learning Databases*; Irvine: Irvine, CA, USA, 1998.
182. Sujatha, K.; Udayarani, V. Chaotic geometric data perturbed and ensemble gradient homomorphic privacy preservation over big healthcare data. *Int. J. Syst. Assur. Eng. Manag.* **2021**, 1–13. [CrossRef]
183. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
184. Cheng, Z.; Jiang, H.; Wang, Y.; Hu, Q.; Yu, J.; Cheng, X. User identity de-anonymization based on attributes. In Proceedings of the International Conference on Wireless Algorithms, Systems, and Applications, Honolulu, HI, USA, 24–26 June 2019; Springer: Cham, Switzerland, 2019; pp. 458–469.
185. Eyupoglu, C.; Aydin, M.A.; Zaim, A.H.; Sertbas, A. An efficient big data anonymization algorithm based on chaos and perturbation techniques. *Entropy* **2018**, *20*, 373. [CrossRef]
186. Luciano, F. Group privacy: A defence and an interpretation. In *Group privacy*; Springer: Cham, Switzerland, 2017; pp. 83–100.
187. Youssef, K.; Qiu, J.; Tan, T.; Cao, G. TargetFinder: A privacy preserving system for locating targets through IoT cameras. *ACM Trans. Internet Things* **2020**, *1*, 1–23.
188. Ruoxuan, W.; Shen, H.; Tian, H. An Improved (k, p, l)-Anonymity Method for Privacy Preserving Collaborative Filtering. In Proceedings of the GLOBECOM 2017–2017 IEEE Global Communications Conference, Singapore, 4–8 December 2017; IEEE: New York, NY, USA, 2017; pp. 1–6.
189. Imran, Z.; Dhou, S.; Judas, J.; Sajun, A.R.; Gomez, B.R.; Hussain, L.A. An IoT System Using Deep Learning to Classify Camera Trap Images on the Edge. *Computers* **2022**, *11*, 13.

190. Mehrdad, J.; Sohrabi, M.K. A Comprehensive Survey on Security Challenges in Different Network Layers in Cloud Computing. *Arch. Comput. Methods Eng.* **2022**, 1–22. [CrossRef]
191. Haiyan, J.; Baumer, E.P.S. Birds of a Feather: Collective Privacy of Online Social Activist Groups. *Comput. Secur.* **2022**, *115*, 102614.
192. Mahdavifar, S.; Deldar, F.; Mahdikhani, H. Personalized Privacy-Preserving Publication of Trajectory Data by Generalization and Distortion of Moving Points. *J. Netw. Syst. Manag.* **2022**, *30*, 1–42. [CrossRef]
193. Dipankar, D.; Chettri, S.K.; Dutta, A.K. Security and Privacy Issues in Internet of Things. In *ICT Analysis and Applications*; Springer: Singapore, 2022; pp. 65–74.
194. Waqas, A.; Nauman, M.; Azam, N. A privacy enhancing model for Internet of Things using three-way decisions and differential privacy. *Comput. Electr. Eng.* **2022**, *100*, 107894.
195. Ti, W.; Zhou, Y.; Ma, H.; Zhang, R. Flexible and Controllable Access Policy Update for Encrypted Data Sharing in the Cloud. *Comput. J.* **2022**. [CrossRef]
196. Luca, C.; Wendzel, S.; Vrhovec, S.; Mileva, A. Security and Privacy Issues of Home Globalization. *IEEE Secur. Priv.* **2022**, *20*, 10–11.
197. Stacey, T.; Baracaldo, N.; Anwar, A.; Steinke, T.; Ludwig, H.; Zhang, R.; Zhou, Y. A hybrid approach to privacy-preserving federated learning. In Proceedings of the 12th ACM Workshop on Artificial Intelligence and Security, London, UK, 15 November 2019; pp. 1–11.
198. Viraaji, M.; Parizi, R.M.; Pouriyeh, S.; Huang, Y.; Dehghantanha, A.; Srivastava, G. A survey on security and privacy of federated learning. *Future Gener. Comput. Syst.* **2021**, *115*, 619–640.
199. Kang, W.; Li, J.; Ding, M.; Ma, C.; Yang, H.H.; Farokhi, F.; Jin, S.; Quek, T.Q.S.; Poor, H.V. Federated learning with differential privacy: Algorithms and performance analysis. *IEEE Trans. Inf. Forensics Secur.* **2020**, *15*, 3454–3469.
200. Yu, B.; Mao, W.; Lv, Y.; Zhang, C.; Xie, Y. A survey on federated learning in data mining. *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.* **2022**, *12*, e1443. [CrossRef]
201. Arif, C.; Pinoli, P.; Gulino, A.; Nanni, L.; Masseroli, M.; Ceri, S. Federated sharing and processing of genomic datasets for tertiary data analysis. *Briefings Bioinform.* **2021**, *22*, bbaa091.
202. Felix, S.; Müller, K.; Samek, W. Clustered federated learning: Model-agnostic distributed multitask optimization under privacy constraints. *IEEE Trans. Neural Netw. Learn. Syst.* **2020**, *32*, 3710–3722.
203. Nader, B.; Mohapatra, P. Vulnerabilities in federated learning. *IEEE Access* **2021**, *9*, 63229–63249.
204. Fan, M.; Haddadi, H.; Katevas, K.; Marin, E.; Perino, D.; Kourtellis, N. PPFL: Enhancing Privacy in Federated Learning with Confidential Computing. *Getmobile Mob. Comput. Commun.* **2022**, *25*, 35–38.