Journal of *Cybersecurity and Privacy*

MDPI

Review

Attribute-Centric and Synthetic Data Based Privacy Preserving Methods: A Systematic Review

Abdul Majeed

Department of Computer Engineering, Gachon University, Seongnam 13120, Republic of Korea; ab09@gachon.ac.kr

Abstract: Anonymization techniques are widely used to make personal data broadly available for analytics/data-mining purposes while preserving the privacy of the personal information enclosed in it. In the past decades, a substantial number of anonymization techniques were developed based on the famous four privacy models such as k-anonymity, ℓ-diversity, t-closeness, and differential privacy. In recent years, there has been an increasing focus on developing attribute-centric anonymization methods, i.e., methods that exploit the properties of the underlying data to be anonymized to improve privacy, utility, and/or computing overheads. In addition, synthetic data are also widely used to preserve privacy (privacy-enhancing technologies), as well as to meet the growing demand for data. To the best of the authors' knowledge, none of the previous studies have covered the distinctive features of attribute-centric anonymization methods and synthetic data based developments. To cover this research gap, this paper summarizes the recent state-of-the-art (SOTA) attribute-centric anonymization methods and synthetic data based developments, along with the experimental details. We report various innovative privacy-enhancing technologies that are used to protect the privacy of personal data enclosed in various forms. We discuss the challenges and the way forward in this line of work to effectively preserve both utility and privacy. This is the first work that systematically covers the recent development in attribute-centric and synthetic-data-based privacy-preserving methods and provides a broader overview of the recent developments in the privacy domain.

Keywords: anonymization; personal data; k-anonymity; ℓ-diversity; t-closeness; differential privacy; synthetic data; privacy enhancing technologies; privacy; attribute-centric anonymization; utility

Citation: Majeed, A. Attribute-Centric and Synthetic Data Based Privacy Preserving Methods: A Systematic Review. *J. Cybersecur. Priv.* **2023**, *3*, 638–661. https://doi.org/10.3390/jcp3030030

Received: 5 July 2023
Revised: 20 August 2023
Accepted: 29 August 2023
Published: 11 September 2023

1. Introduction

Recently, personal data have become an economically desirable resource, and they offer valuable knowledge that can influence science and advance societies. Personal data can be used to improve real-world data-driven services such as improved and low-cost healthcare, better recommendations by utilizing heterogeneous data, and increase the precision of navigation information. However, personal data often encompass sensitive information and require privacy preservation in processing and dissemination [1]. According to a survey [2], user privacy can be compromised by using different attributes present in the data even if some unique information is removed from the data. The survey findings are summarized in Figure 1, where three combinations of attributes have a distinct impact on users' re-identification in a dataset.

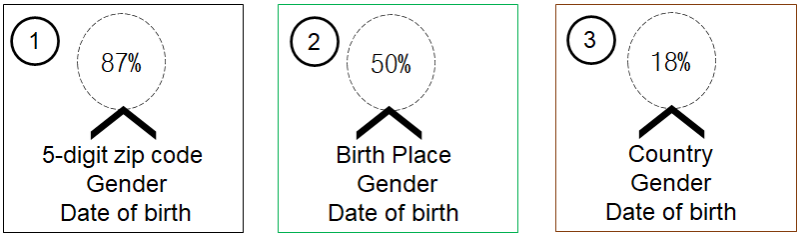


Figure 1. Impact of the user attributes' items on the users' identification in a dataset.

As shown in Figure 1, a substantial number of users can be identified from the datasets, and many sophisticated techniques are required to address the privacy issues. There are three re-identification scenarios in Figure 1. In the first scenario, the much higher re-identification is due to the zip code, which precisely locates many individuals. In the second and third cases, the re-identification is lower due to the generalization of residence. These results highlight the need for effective privacy preservation in data analysis. The privacy preservation of data while offering higher knowledge is a global problem, and many techniques were developed to address this trade-off [3]. There are three important challenges while handling personal data encompassing basic as well as sensitive information about individuals.

- Resolution of privacy versus utility trade-off: How to safeguard user privacy while still allowing data miners/analysts to maximally extract the enclosed knowledge from the personal data.
- Preventing the misuse of personal data: How to enable fair and impartial decision making concerning real-world entities, and restricting target profiling (or discrimination about a minor community).
- Enhancing the quality of personal data for the well-being of societies: How to improve the quality of the data when they are either small or of low quality to enable better data mining and decision making.

It is important to note that all of the above-cited challenges are closely related to each other. For example, if the trade-off between privacy and utility is effectively resolved, the misuse of personal data can be restrained, and the personal data can be used as intended. Privacy disclosures occur when either inappropriate anonymization models are used in anonymization or the underlying dataset to be anonymized is of poor quality. A dataset is considered poor quality from the utilization perspective when the intended knowledge cannot be extracted from it with minimal effort. For example, medical data can be regarded as poor quality when a medical student interested in exploring the disease connection with various demographics cannot find relevant information from the data. Similarly, if the data do not encompass diverse and complete information, they are also referred to as poor quality. The quality of the data from within the utilization perspective varies from application to application and case to case. In contrast, a dataset that is complete, diverse, and representative of the problem under investigation is considered to be of rich quality. Therefore, the misuse of personal data can be prevented by applying careful anonymization and curating more data to improve the bad parts of the data (e.g., the data concerning minor population groups where the risk of discrimination is high). However, the optimization of the privacy–utility trade-off, the prevention of the misuse of data, and the improvement in data quality are very challenging to achieve simultaneously. Most of the existing methods often pay attention to one metric and compromise the others. It is vital to achieve all three objectives simultaneously by paying careful attention to data properties, data pre-processing, and anonymization.

To address the challenges of the privacy versus utility trade-off, a variety of anonymization methods were developed. The majority of the developed methods are derived from the four methods, such as k-anonymity [4], ℓ-diversity [5], t-closeness [6], and differential privacy [7]. Apart from these mainstream solutions, many enhancements of these methods were proposed to optimize the privacy and utility in personal data handling [8–10]. To prevent the misuse of personal data, techniques such as synthetic data, legal measures, consents, encryption, and other privacy-enhancing technologies, etc., are used [11–13]. Recently, some work has begun on data quality enhancement to improve decision-making and to solve industrial problems [14,15]. The data-tailored practices are vital to improving the quality of people's lives with improved decision-making. Unfortunately, there is a serious lack of data-centric methods in the privacy domain, and only a few methods have considered the properties of data during their anonymization. In addition, none of the previous studies have highlighted the need for data- or attribute-centric methods in the privacy domain. The major contributions of this paper are summarized below.

- We discuss the major research tracks in the information privacy domain with a specific emphasis on attribute-centric anonymization methods that were recently developed to address the privacy versus utility trade-off.
- We discuss synthetic data generation methods and the role of synthetic data as a privacy-enhancing technology, as well as a data quality enhancement technology.
- We highlight the various privacy-enhancing technologies that are widely used to preserve the privacy of the personal data enclosed in heterogeneous formats.
- We suggest promising research tracks for future work that require the immediate attention of the privacy community amid the rapid rise in digitization.
- To the best of the authors' knowledge, this is the first work that discusses two feature-oriented privacy-enhancing technologies (i.e., attribute-centric and synthetic data) from a much broader perspective. We hope to provide a solid foundation for future research by making a timely contribution to this line of work.

There exist plenty of metrics for comparing synthetic data generation quality from multiple perspectives. Emam [16] suggested seven different ways to gauge the quality of synthetic data in real-world cases. Mostly, the distribution and structural similarities are analyzed to check the closeness between the synthetic and real data. In some cases, the utility of the synthetic data is evaluated with respect to the target application (e.g., health care) for which the data are being curated [17]. A comprehensive list of metrics that can be used to evaluate the quality of synthetic data can be learned from Alvaro et al. [18]. In some cases, the evaluation of synthetic data is performed from the perspectives of both privacy and utility [19]. In most of the existing papers, the comparison regarding the quality of the synthetic data is made with real data or the previous algorithms that were proposed for similar tasks (e.g., synthetic data generation). Similarly, the requirements for the attribute-centric methods are to exploit any useful knowledge that can either lessen the computing complexity of privacy models or assist in effectively resolving the trade-offs between privacy and utility. Only a limited number of papers have been proposed in this line of work that exploit valuable knowledge about real data to improve privacy/utility results [20,21]. In [20], the authors considered the preferences of users along with their data at anonymization time to improve data quality in data publishing scenarios. In [21], the authors devised a method to rank the attribute first, and then, KNN clustering was applied to enable the anonymity of the datasets lacking diversity. The proposed method reduces the overgeneralization issues, thereby achieving privacy protection for all attributes. The major requirement for the attribute-centric methods is any piece of knowledge regarding real data composition or attributes' values that can ease the anonymity process. The empirical and formal comparison can be made to contrast these requirements and accomplishments towards these requirements in personal data handling scenarios [22,23]. The use cases discussed in these studies are privacy protection when a database encompasses more than one SA and permutation of the dataset for fulfilling k-anonymity criteria, respectively. Feature-based comparisons between traditional and attribute-centric methods can also be performed to highlight the novel aspects of the attribute-centric methods.

There exists a plethora of survey articles on privacy-preserving data mining (PPDM) for a variety of applications such as cloud computing, location-based services, e-health, recommender systems, transport data, and internet of things [24–28]. In PPDM, privacy is ensured by applying the anonymization method while minimally changing the semantics of the data [25]. The objective of PPDM is to maximize utility with considerable privacy guarantees. In PPDM, the data are not released publicly in some cases, and instead, queries are executed in a privacy-preserved way. On the other hand, PPDP explores ways to publish anonymized data either in full or partial form so that they can be used by relevant information consumers [29]. This work presents the latest developments in the PPDP area and has two main differences from the existing surveys. First, we identify and discuss the data generation methods along with the privacy guarantees (e.g., GAN and DP are simultaneously used or GAN is used first to curate the data, which are later anonymized via DP or general anonymity solutions) to compensate for the deficiency in

the data. The privacy-preserved synthetic data can also be used to train AI models while providing privacy guarantees against some threats such as membership inference attacks. Second, we describe the attribute-centric methods that extract some knowledge from the data composition to ease the anonymization process. To the best of our knowledge, both of these methods have not been thoroughly covered in the previous surveys. To cover this research gap, our survey can provide a solid foundation for future studies in these lines of work (e.g., synthetic-data-based privacy-preserving methods and attribute-centric anonymization methods). Finally, we provide sufficient experimental details of the previous studies concerning these two types of methods, which can help researchers quickly grasp the research status and further advance the status of these developments.

The rest of this paper is structured as follows. Section 2 discusses the process of privacy-preserving data publishing (PPDP) and outlines eleven major research tracks of PPDP. Section 3 presents in-depth details of the state-of-the-art (SOTA) attribute-centric anonymization methods that were recently proposed to strike the balance between privacy and utility. Section 4 discusses the recent SOTA synthetic-data-based methods that were recently proposed to fulfill privacy and data needs. Section 5 analyzes the privacy-enhancing technologies that were developed to enhance the protection against present-day privacy threats. Section 6 uncovers the future research and development directions to foster further developments in the information privacy domain. We conclude this paper in Section 7.

2. Privacy Preserving Data Publishing and Major Research Tracks

2.1. Privacy Preserving Data Publishing

In this section, we explain the privacy-preserving data publishing (PPDP) process and outline its major steps. Furthermore, we also discuss the major research tracks of the PPDP. A conceptual overview of PPDP to foster the secondary use of personal data is demonstrated in Figure 2.

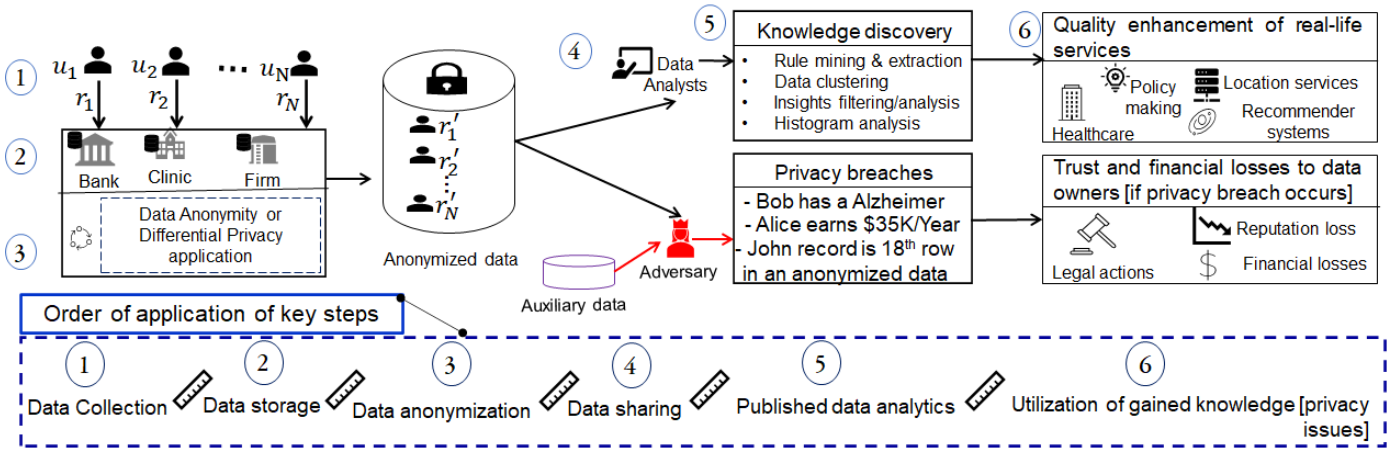


Figure 2. Schematic of privacy-preserving data publishing (PPDP) for the secondary use of data.

The whole PPDP process encompasses five key steps. In the first step, the data are collected from the relevant people. The collected data can be of any type such as demographics, information about monthly earnings, disease contracted, etc. In the second step, the data are stored, and pre-processed (the pre-processing includes data cleaning and quality enhancement (removing redundant records, removing outliers, handling missing values, and giving simplified structure to the data)) in the data owners' environments. The data owners can be hospitals, banks, insurance companies, social network service providers, etc. Later, the data are anonymized with the help of generalization hierarchies or the noise addition of DP. In the fourth step, the data are outsourced to data analysts or miners to extract the enclosed knowledge. The data are published either directly by the data owners or with the help of third parties. In the fifth step, the published data are analyzed by analysts or data miners with the help of advanced data mining or machine

learning tools. In the last step, the knowledge gained from the published data is used to improve real-world services such as healthcare, smart cities, recommendation systems, navigation services, and other data-driven services [30,31]. Data sharing has become a routine matter amid the rapid rise in digital developments. In addition, data sharing is imperative to improve the quality of data-driven applications/services in the modern era.

Although a substantial number of studies were proposed to address privacy issues in PPDP, there are still two major trade-offs that are hard to achieve. We present both trade-offs in Figure 3. In the first trade-off, the semantics of the original data need to be preserved [32]. In contrast, for the second trade-off, it is desirable to prevent minor values from dilution during the anonymization [33]. Despite many developments, both these trade-offs are quite challenging to achieve in the PPDP scenarios. The effective resolution of the privacy–utility trade-off is hard to achieve owing to strict privacy parameters to be applied during anonymization, poor quality of the underlying data, strict privacy/utility guarantees, and lack of evaluation metrics. In some cases, the selection of the optimal level/ϵ to perform data generalization/transformation leads to the solution of one metric, i.e., either utility or privacy [34]. Similarly, the privacy–equity trade-off cannot be achieved owing to the poor quality (e.g., data imbalance) of the original data. For example, if there are fewer representations of some minor groups in the real data, the anonymization method further reduces the representation, leading to fewer benefits for some minority groups. Most anonymization methods do not improve data quality before anonymization, and therefore, it is very hard to achieve the trade-off between privacy and equity. In the modern era, it was suggested to improve data quality to lower the data-driven threats [14]. In conclusion, there are many obvious reasons (e.g., poor data quality, inappropriate anonymization, data owner requirement, privacy and utility objectives, data publishing goals, wrong data collection processes, and malicious entities present in the PPDP process) that make these trade-offs hard to achieve. All three trade-offs interact with each other because the data portion, which contains more information for analytics/mining, may require better privacy preservation. However, the data owners can make choices regarding these trade-offs to be resolved depending upon the goals of data outsourcing. It is worth noting that the quantification of these trade-offs depends on evaluation metrics, and therefore, the selection of optimal metrics is desirable to quantify them accurately in various applications.

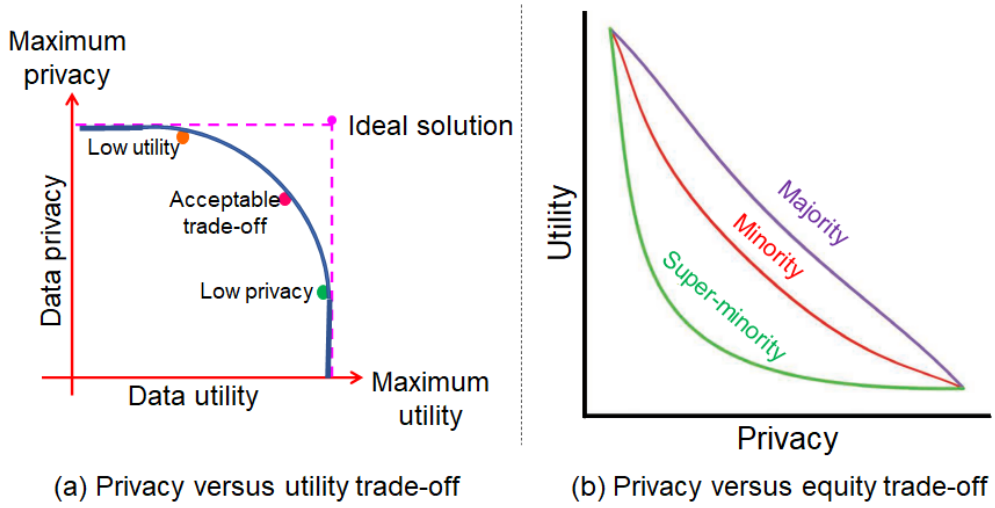


Figure 3. Overview of two major trade-offs in the privacy preservation domain.

2.2. Privacy Models

Next, we define four well-known and state-of-the-art privacy models used for privacy-preserving data publishing.

Definition 1 (k-anonymity)**.** *In this model, T (relational data) is divided into small, non-overlapping classes of size k where all users' QIDs in a class are the same. For example, if $k = 2$, there must be at least two users in a class. The probability of re-identifying someone from the anonymized data via the k-anonymity model is $1/k$.*

Definition 2 (ℓ-diversity)**.** *This model solely focuses on the SA values present in T, ℓ-diversity specifically ensures that each class has ℓ distinct SA values. When $\ell = 2$, every class must encompass at least two different SA values. The probability of re-identifying someone's SA from the data via the ℓ-diversity model is $1/\ell$.*

Definition 3 (t-closeness)**.** *Similar to ℓ-diversity, t-closeness also focuses on SA values present in T. A T is t-close only when the distribution of SAs in each class and all of T is $\leq t$, where t is a threshold. In simple terms, SA values are fairly allocated to each class under the influence of t. T is said to be t-close if every class is t-close.*

Definition 4 (Differential privacy)**.** *DP anonymizes data by adding noise and using randomization operations. DP and its enhancements are called semantic methods. In DP, $\mathcal{F}$ (a randomized function) guarantees ϵ-DP if $\forall$ raw datasets, T_1 and T_2, differ by at most one record, and $\forall$ $\mathcal{G} \subseteq Range(\mathcal{F})$,*

$$Pr[\mathcal{F}(T_1) \in \mathcal{G}] \leq exp(\epsilon) \times Pr[\mathcal{F}(T_2) \in \mathcal{G}] \quad (1)$$

DP can be satisfied via exponential and Laplace mechanisms considering the nature of the data. Although t-closeness is an enhancement of k-anonymity and ℓ-diversity models, imbalanced distributions of SA can severely affect the performance of the t-closeness model. Furthermore, the utility loss from the t-closeness model is significantly higher compared to the first two models (e.g., k-anonymity and ℓ-diversity). In all three models, generalization or suppression is mostly used to anonymize the data [35]. For example, if age = 55, it will likely be generalized in one of two ways: $\geq$50 or 50–60. Generalization is performed with the help of pre-built generalization taxonomies for each QID. Suppression is a special case of generalization that fully hides values of attributes with an asterisk (*). DP has become a state-of-the-art model for privacy preservation in dynamic and static scenarios. Although DP provides much higher privacy guarantees, the utility of the resulting data is lower in most cases.

2.3. Major Research Tracks in Privacy Preserving Data Publishing

A variety of techniques were developed to safeguard user privacy in personal data handling. To this end, famous solutions are encryption, masking, watermarking, secret sharing, secure multiparty computation, anonymization, and secure enclaves [36,37]. Despite other solutions, anonymization is one of the most widely used tools for preserving the privacy of user data owing to the least computing complexity, the ease in employment, and the conceptual simplicity. Anonymization technology has significantly advanced from multiple perspectives, and many independent tracks exist thus far. Figure 4 systematically presents the advancements in the privacy domain. The anonymization concept started with the syntactic privacy methods and advanced to many other latest methods, as shown in Figure 4. For example, the syntactic methods yielded poor utility and were improved by the clustering-based methods. The DP-based methods have shown remarkable achievements in static and dynamic scenarios. For example, DP-based methods were implemented widely in different domains such as federated learning, IoT, industrial IoT, and cloud computing environments for privacy preservation. Many AI-based methods were also used to improve many critical parts of the traditional anonymity methods and to improve the privacy and utility results [38]. Some methods were developed to counter a specific privacy threat or to preserve privacy in a specific data style (i.e., table, graph, matrix, text, trace, etc.). In some cases, more than one method was jointly used to protect the privacy of user data. To the best of our knowledge, detailed tracks and famous methods under each track, in

particular, attribute-centric anonymization and synthetic-data-based methods, have not been discussed in previous studies.

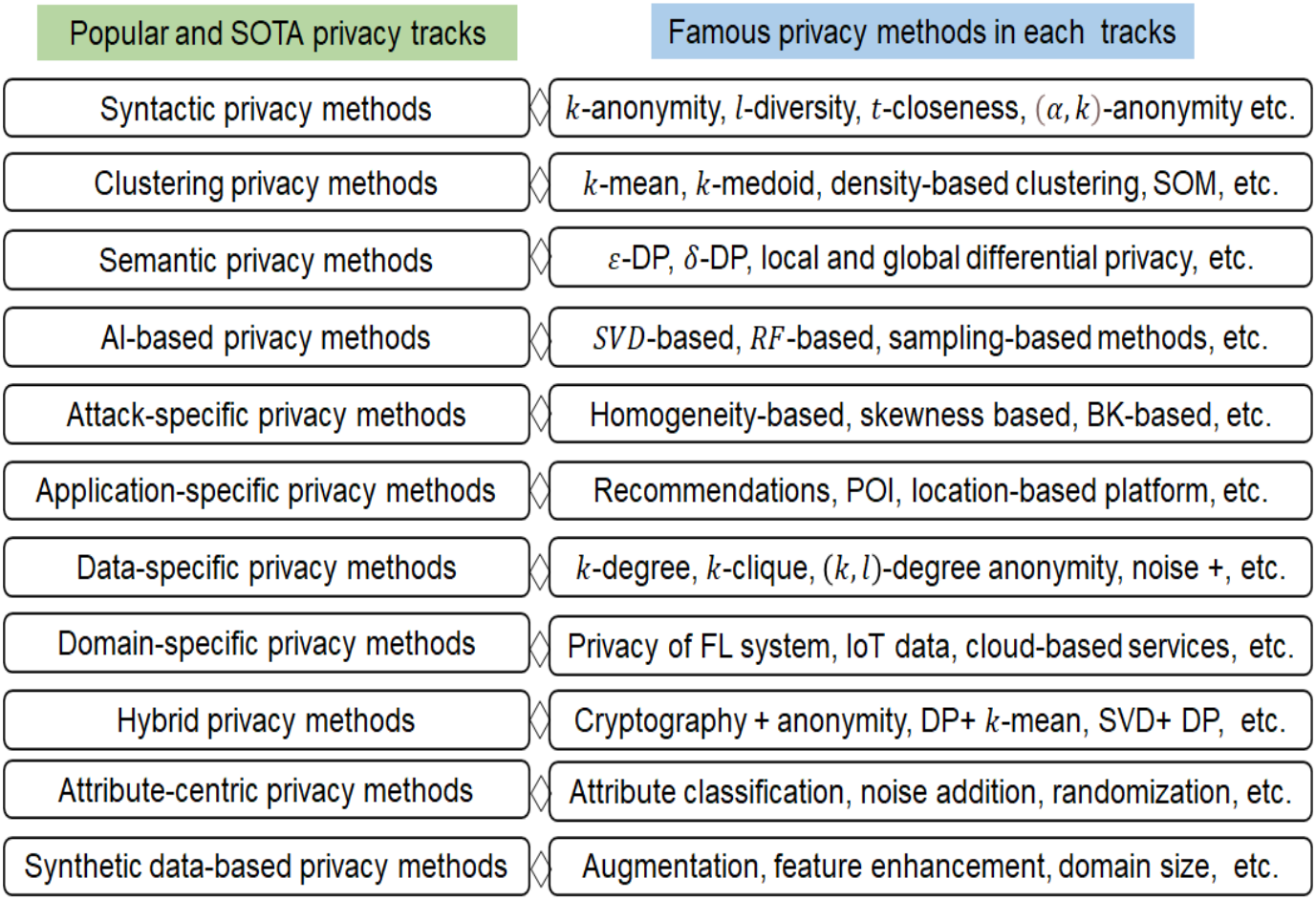


Figure 4. Schematic of famous and SOTA privacy tracks and methods developed from 2002∼2023.

Figure 5 demonstrates the data sources used in this study. We extracted the relevant information from popular sources with the help of relevant search strings.

The distinctive features of the attribute-centric methods are to extract valuable knowledge to the extent possible about the underlying data and to reduce anonymity operations. In contrast, the unique features of synthetic-data-based methods are to mimic the properties of real data as much as possible and curate close copies of real data. Both these methods are developed to meet the growing demands of both privacy and utility in the big-data- and AI-driven era. The attribute-centric methods have a close relationship with the three famous syntactic methods (e.g., k-anonymity, ℓ-diversity, and t-closeness). In contrast, synthetic-data-based developments have a strong connection with differential privacy-based methods.

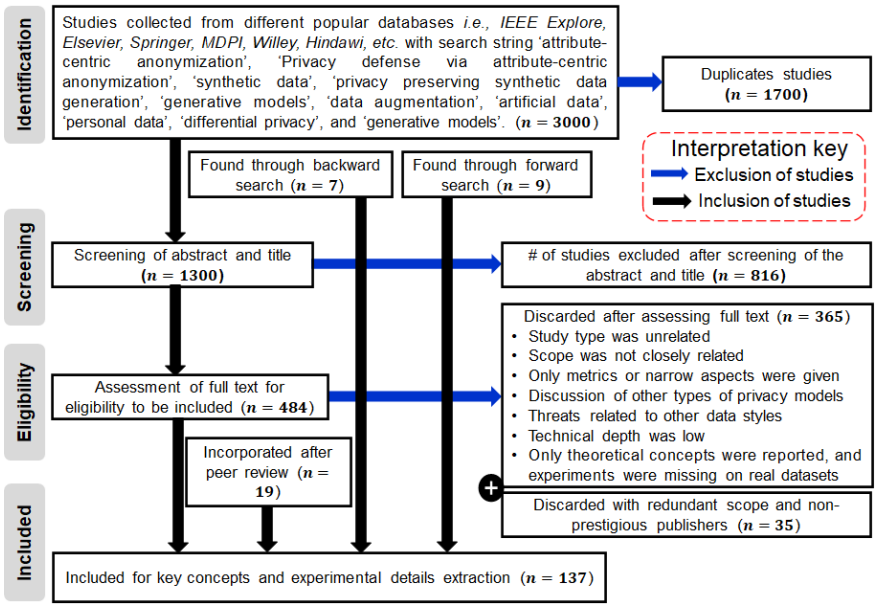


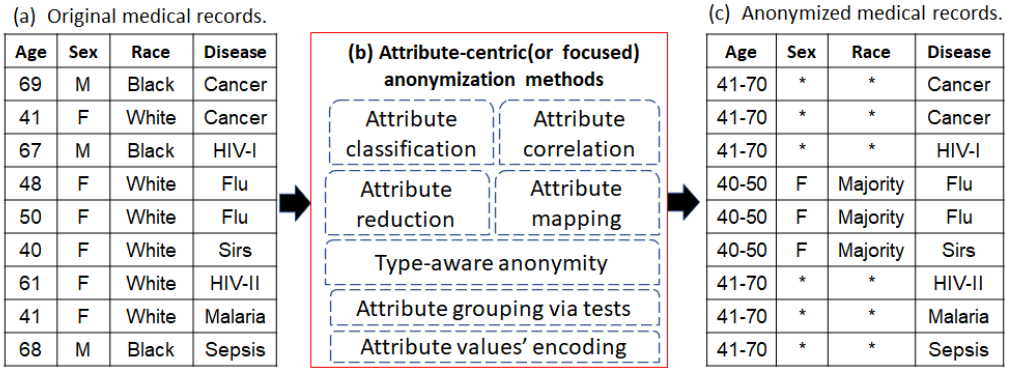
Figure 5. Overview of the data collection process used in this study.

2.4. Attribute-Centric and Synthetic-Data-Based Privacy-Preserving Methods

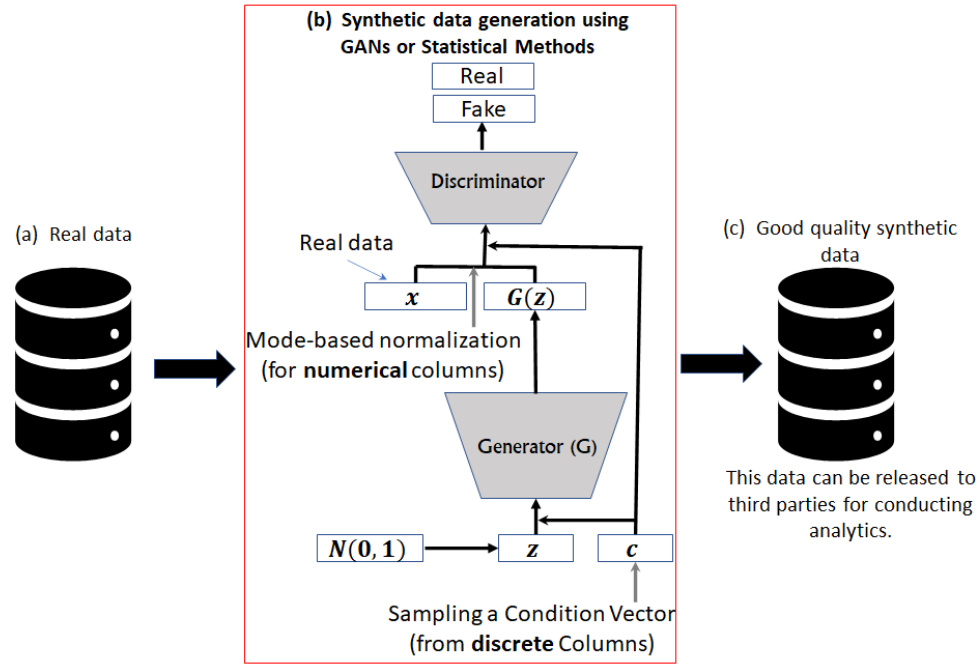
In recent years, there has been an increasing focus on developing attribute-centric and synthetic-data-based privacy-preserving solutions. In the former category, the characteristics of the underlying data are exploited to strike the balance between privacy and utility [39]. In the latter category, virtual samples are generated to augment the data quality without compromising privacy [40]. In this paper, our main focus is on attribute-centric and synthetic-data-based privacy-preserving solutions that remained unexplored in the previous research. Figure 6 demonstrates the conceptual overview of both these methods. Referring to Figure 6a, the tabular data on the right side are real data, and the left side shows the anonymized data. In these data, the last column is a sensitive attribute (SA), and the first three columns are quasi-identifiers (QIDs).

Referring to Figure 6a, the detailed information concerning the composition of the data is exploited to ease the anonymity process. For example, by classifying data and exploiting the co-relations between attributes, privacy and utility can be effectively preserved [41]. By exploiting information concerning data, computing power can also be saved, and only the relevant parts can be anonymized [42]. Referring to Figure 6a, the conditional generative adversarial (CTGAN) model is employed to create synthetic data by mimicking the properties of real data. In the CTGAN model, the important component is the conditional vector that guarantees the data generation of the best quality. Recently, many variants of GAN models were developed to curate data of diverse modalities (e.g., time series, images, tabular data, graphs, etc.) for different use cases [43–45]. Both these methods are very recent and significantly contribute to data outsourcing while preserving data privacy. In recent years, many privacy-preserving methods were developed to securely publish personal data. Hongbin et al. [46] proposed a multidimensional data aggregation and privacy-preserving scheme based on the federated learning concept. The proposed scheme can be used in industrial IoT domains to restrict privacy breaches. Paul et al. [47] discussed the security and privacy concerns in the healthcare sector due to the rapid rise in digital technologies. Muneeswari et al. [48] developed a privacy-preserving framework for self-diagnosis based on IoT devices. The proposed framework utilizes patient records and performs analysis without losing guarantees of privacy. Xie et al. [49] proposed a blockchain-based framework for privacy preservation in IoT scenarios. Liu et al. [50] proposed a novel method by amalgamating machine learning techniques and conditional probability distributions to effectively resolve the privacy–utility trade-off. The proposed method can be used to preserve privacy in single as well as multiple SA

scenarios. Hewage et al. [51] discussed the PPDSM and PPDM methods used for privacy protection. The authors analyzed the strengths and weaknesses of the existing methods and stressed the need for more methods in the PPDSM domain in the future. Terziyan et al. [52] proposed a method for privacy preservation in the image data. The authors used privacy protection methods with (convolutional, variational) autoencoders to preserve the privacy of the image data. Qin et al. [53] developed a scheme to trace virus-infected people without exposing their privacy. The authors proposed a new IoT data management architecture and applied various perturbation techniques to effectively preserve privacy and availability in the cloud environments. Kumuthini et al. [54] proposed two methods (i.e., archive and data commons) for genomics data sharing. The authors highlighted the challenges involved in data sharing and discussed privacy and security concerns. Yang et al. [55] proposed a method for privacy preservation in query execution. The proposed method protects the exposure of personal information in query-based systems and answers queries without breaching the confidentiality of the underlying data. Table 1 presents the summary and comparisons of the above-cited SOTA approaches from a broader perspective.



(a) Conceptual overview of the attribute-centric privacy methods.



(b) Conceptual overview of the synthetic data-based privacy methods.

Figure 6. Conceptual overview of attribute-centric and synthetic-data-based privacy methods.

Table 1. Summary and comparisons of most recent and SOTA privacy-preserving methods.

Proposed Approach	Challenge (s)	Benefits	Contribution (s)	Reference
FL-based micro aggregation	Privacy–utility trade-off	Privacy in IIoT domains	Strong defense against recent privacy threats	Hongbin et al. [46]
Anonymization based techniques	Security and privacy challenges	Strong privacy of health data	Uncover privacy and security needs	Paul et al. [47]
Similarity-based analysis	Privacy in medical diagnosis	Strong privacy in cloud setting	A low-cost privacy-preserving framework	Muneeswari et al. [48]
Blockchain-based system	Privacy protection of real data	Privacy of sensors data	A robust defense mechanism	Xie et al. [49]
ML + CPD method	Privacy–utility trade-off	Application in SSA and MSA scenarios	A hybrid method for PPDP	Liu et al. [50]
PPDSM and PPDM methods	Privacy in data mining	Protection of confidential data	Analysis of various methods	Hewage et al. [51]
CV autoencoders	Identity protection in image data	Privacy of image data	Low cost perturbation methods	Terziyan et al. [52]
Blockchain-based system	Privacy of virus-infected users	Virus control and mitigation	Effective approach for privacy protection	Qin et al. [53]
Archive and data commons	Disclosure of sensitive data	Privacy of genomics data	Proposed ways to address privacy issues	Kumuthini et al. [54]
PPQE method	Privacy of confidential data	Responsible use of data	Reliable perturbation methods	Yang et al. [55]
DPView system	High-dimensional data handling	Better utility of SD	Data curation with privacy	Lin et al. [56]
DP model	Missing values handling	Informative analysis of COVID-19 data	Curating better data	Sei et al. [57]

Abbreviations: ML = machine learning, CPD = conditional probability distribution, SSA = single SA, MSA = multiple SA, CV = convolutional and variational, PPQE = privacy-preserving query execution, IIoT = industrial IoT, DP = differential privacy.

3. Discussion on Attribute-Centric Privacy-Preserving Methods

Recently, many attribute-centric privacy methods were developed to secure personal data from adversaries while sustaining higher utility in anonymized data. These methods exploit characteristics of the underlying data to be anonymized and perform the required operation to accomplish the conflicting goals. In some cases, attribute-centric privacy methods were used to restrict changes in the anonymization process [58,59]. In some cases, these methods were used to clean the data from the perspective of outliers, missing values, and/or redundant records to increase data utility [60]. Table 2 presents the technical details and comparisons of recently developed attribute-centric privacy methods. Each method listed in Table 2 has used a different mechanism to perform anonymization of the real data. The commonly used mechanisms in these studies are micro-aggregation, clustering, differential privacy, feature-aware anonymization, fuzzy *c*-means clustering, diverse grouping, chaos and perturbations, similarity-aware clustering, and generation of fair equivalence classes.

Table 2. Summary and comparisons of the most recent attribute-centric privacy methods.

Techniques Used	Objective (s)	Application Area	Study Type	Data Type	Reference
Fixed intervals + IDs generation	Privacy–utility trade-off	Healthcare	Technical	Real	Majeed et al. [61]
Fixed intervals+ Improved ℓ-diversity	Privacy–utility trade-off	Healthcare	Technical	Real	Onesimu et al. [62]
Hybrid schemes	Privacy–utility enhancement	Medical data	Technical	Real	Hui et al. [63]
Uncertainty + deviation	Privacy–utility enhancement	General scenarios	Technical	Real	Khan et al. [64]
DP + tree model	Data utility and patient's privacy	Medical data	Technical	Real	Zhang et al. [65]
Three syntactic models	Privacy–utility enhancement	General scenarios	Technical	Real	Sadhya et al. [66]
Feature selection + anonymization	Data utility enhancement	General scenarios	Technical	Real	Srijayanthi et al. [67]
Mondrian approach	Data utility enhancement	General scenarios	Technical	Real	Canbay et al. [68]
Analytical approach	Privacy–utility enhancement	Smart health data	Technical	Real	Arca et al. [69]
k-CMVM and Constrained-CMVM	Utility enhancement	General scenarios	Technical	Real	Zouinina et al. [70]
Micro-aggregation approach	Privacy enhancement	dynamic data release	Technical	Real	Yan et al. [71]
Util-MA approach	Reduction in Iloss	Machine learning applications	Technical	Real and synthetic	Lee et al. [72]
Grid clustering + DP	Query accuracy	Location data sharing	Technical	Real and synthetic	Yan et al. [73]
AFBSO + WOA	Privacy and utility enhancement	Healthcare data	Technical	Synthetic	Thanga et al. [74]
GM-FBO algorithm	Preserving privacy of SHD	Cloud computing	Technical	Real	Anand et al. [75]
CGBFO-GC algorithm	Multi-privacy objectives	Cloud computing	Technical	Real	Anand et al. [76]
OAN model	Compute cost reduction	General scenarios	Theoretical	Synthetic	Canbay et al. [77]
Clustering method	Privacy and utility enhancement	IoT environments	Technical	Real	Onesimu et al. [78]
Fuzzy clustering	Privacy and utility enhancement	Industrial IoT	Technical	Real	Xie et al. [79]
IDEA method	Effectively preserving utility.	General scenarios	Technical	Real	Yang et al. [80]
(a, k)-anonymous	Better privacy and data quality	IoT-based healthcare	Technical	Real	Li et al. [81]
BL approach	Data Security	Medical healthcare	Technical	Real	Altameem et al. [82]
Clustering approach	Data Security and utility	General scenarios	Technical	Real	Nayahi et al. [83]
DHkmeans-ℓ-diversity	SA privacy protection	Big data era	Technical	Real	Ashkouti et al. [84]
δ-value approach	Data mining	Information retrieval	Technical	Real	Solanki et al. [85]
CAP approach	PPDP and PPDM	Knowledge discovery and mining	Technical	Real	Eyupoglu et al. [86]

Abbreviations: ILoss = information loss, AFBSO = adaptive fractional brain storm optimization, WOA = whale optimization algorithm, GM-FBO=Gaussian mutation-based firebug optimization, SHD = sensitive healthcare data, CGBFO-GC = Chaotic chemotaxis and Gaussian mutation-based bacterial foraging optimization with a genetic crossover operation, IDEA=Incomplete Data strEam Anonymization, BL = backpropagation learning, SA = sensitive attributes, CAP=chaos and perturbation, PPDM = privacy-preserving data mining.

4. Discussion on Synthetic Data-Based Privacy Methods

Recently, synthetic data have been widely used to preserve the privacy of users while fulfilling the data analytics needs [87,88]. In recent years, a large amount of data has been needed to train AI models for enhancing accuracy, particularly in the healthcare domain [89]. To this end, synthetic data can be employed to increase the number of samples in training data. Synthetic data have become a SOTA technology with a wide range of practical applications in diverse sectors. Figure 7 presents the practical uses of synthetic data in the modern era. From Figure 7, it can be seen that synthetic data are widely used as a replacement for real data in diverse sectors. The application of synthetic data in the healthcare sector is also steadily increasing with time [90]. Figueira et al. [18] recently presented a comprehensive survey on different tools that can be used to generate synthetic data (in a tabular form), particularly generative adversarial network (GAN) architectures. Gonzales et al. [91] recently discussed the innovative uses of synthetic data in healthcare. Based on these major developments, it is fair to say that synthetic data are one of the leading technologies of the future. This paper focuses on privacy preservation, and therefore, we demonstrate synthetic-data-based privacy methods. This topic has attracted the researchers' attention in recent years, and many notable developments have been made [92].

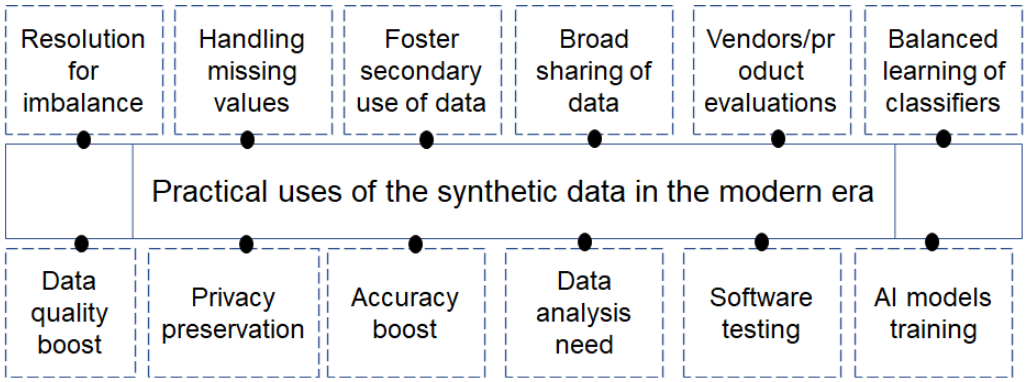


Figure 7. Practical uses of synthetic data in the modern era (e.g., the year 2023 and beyond).

Table 3 presents the comparison of SOTA synthetic-data-based privacy methods. Specifically, we compare each method on five grounds (i.e., techniques used in the study, objective accomplished with the study, type of the study (technical or theoretical), and data type used in each study). Furthermore, we considered the most recent studies to provide a fresh analysis of the published literature. The analysis presented in Table 3 can pave the way to understanding the latest developments in synthetic-data-based privacy methods. Each method listed in Table 2 has used a different mechanism to generate synthetic data of good quality to either preserve the privacy of personal data or to fulfill data requirements/needs. The commonly used mechanisms in these studies are differential privacy, probabilistic modeling, variational autoencoders, generative adversarial networks, neural networks, generative models, pipelines including differential privacy, clustering methods, and transformer models. In conclusion, most of these methods have used basic generative adversarial networks (GANs) and their enhancement to generate data of diverse types. In these methods, privacy is achieved using two ways: (i) including DP or any other privacy models with the generative models, and (ii) generating synthetic data and applying anonymization to guarantee privacy. The synthetic-data-based methods are widely used to accomplish five key requirements in the modern era: (i) privacy preservation of personal data, (ii) testing of software/products, (iii) data governance, (iv) data sharing at a scale, and (v) training machine learning classifiers [93].

Table 3. Summary and comparisons of the most recent synthetic data-based privacy methods.

Techniques Used	Objective (s)	Application Area	Study Type	Data Type	Reference
DP+ MERF approach	produce tabular and image data with privacy guarantees	General scenarios	Technical	Real	Harder et al. [94]
DP-HFlow method	Privacy protection in data sharing	General Scenarios	Technical	Real	Lee et al. [95]
Probabilistic modeling	Anonymized synthetic data sharing	Open science	Technical	Real and synthetic	Jälkö et al. [96]
HealthGAN model	Better data analysis	Education and research	Technical	Synthetic	Yale et al. [97]
GAN+ XAI	High quality SD generation	Health data	Technical	Real	Lenatti et al. [98]
HealthGAN model	Capturing trends from TSD	Medical domain	Technical	Real and synthetic	Bhanot et al. [99]
VGAE model	Yield artificial trajectories with PPP	Electronic health	Technical	Synthetic	Nikolentzos et al. [100]
VITALISE model	Compliance-based data use	Health and well-being domain	Technical	Real	Hernandez et al. [101]
GAN model	Reduce risk of SA disclosure	Fitness related	Technical	Real	Kuo et al. [102]
SDG framework	Privacy preservation and CP	Medical domain	Technical	Real	Rodriguez et al. [103]
dsSynthetic package	Data harmonization	General scenarios	Technical	Real	Banerjee et al. [104]
STSG approach	Privacy guarantees in TSD	General scenarios	Theoretical	Real	Larrea et al. [105]
pGAN model	Privacy guarantees in EHR	Medical domain	Technical	Real	Venugopal et al. [106]
VAE model	Fixation of bias and privacy	Medical domain	Technical	Real and synthetic	Yoshikawa et al. [107]
Neural-Prophet model	Maintaining validity of MD	Medical systems	Technical	Real	Hyun et al. [108]
Transformer models	Accurate clinical predictions with privacy guarantees	healthcare	Technical	Real	Zhang et al. [109]
HealthGAN model	Privacy, utility, and resemblance	Healthcare domain	Technical	Real	Yale et al. [110]
GANs models	Data augmentation	General scenarios	Technical	Real	Narteni et al. [111]
GAN model	Control on various privacy risks	Big data apps	Theoretical	Real	Raveendran et al. [112]
MC-GEN model	Privacy guarantees in classification tasks	ML applications	Technical	Real	Li et al. [113]
PPEA model	Better utility of data	Distributed environments	Technical	Real	Shahani et al. [114]
DP+ GAN	Higher privacy guarantees	Industrial IoT	Technical	Real	Hindistan et al. [115]
(ϵ, δ) -ULDP	Strong privacy protection	General scenarios	Technical	Real, synthetic	Zhang et al. [116]
Fed Select Framework	Strong privacy guarantees in FL	IoMT settings	Technical	Real	Nair et al. [117]
LGAN + DP	Privacy-utility trade-off	ML applications	Technical	Real	Zhang et al. [118]
HT-Fed-GAN model	Privacy–utility trade-off	machine learning tasks	Technical	Real	Duan et al. [119]

Abbreviations: SD = synthetic data, TSD = time series data, VGAE = variational graph autoencoder, PPP = patient privacy preservation, pGAN = privacy-preserving generative adversarial network, EHR = electronic health records, VAE = variational Autoencoder, PPEA = privacy-preserving endpoint aggregation, FL = federated learning, ML = machine learning.

Recently, synthetic data (SD) have become an indispensable component for AI applications, and therefore, many generative models were recently developed to compensate for the deficiency in good data [120,121]. The SD generated with the generative models has lots of applications in diverse fields such as medical image analysis, fault diagnosis, language models, image recognition, and medical diagnosis, to name a few. Furthermore, some companies are sharing SD rather than real data to overcome privacy concerns. Figure 8 presents SD-based developments along with the experimental details. We classify the SD generation process into four key steps, as marked in by the color yellow. As shown in Figure 8, the real data are mostly limited to the data owners' environments, and therefore, generative models of a suitable type can be moved closed to that data to produce synthetic data. The structure of each generative model is different, and therefore, a suitable model is chosen based on the data type. For instance, to generate tabular data, Table-GAN can be used. In contrast, DCGAN is suitable for generating synthetic images. Similarly, for the time-series data, a different GAN model can be used. After training GAN models by connecting with real data, the SD can be obtained in diverse formats depending upon the application. Later, the SD can be used for different purposes, as shown in Figure 8(4). In recent years, SD has been widely used for data augmentation purposes while building classifiers. With the advent of the DC-AI concept, SD has become an enabling technology for AI applications. In the coming years, SD will be used as an integral component for medical applications, where access to real data is often prohibited owing to privacy concerns.

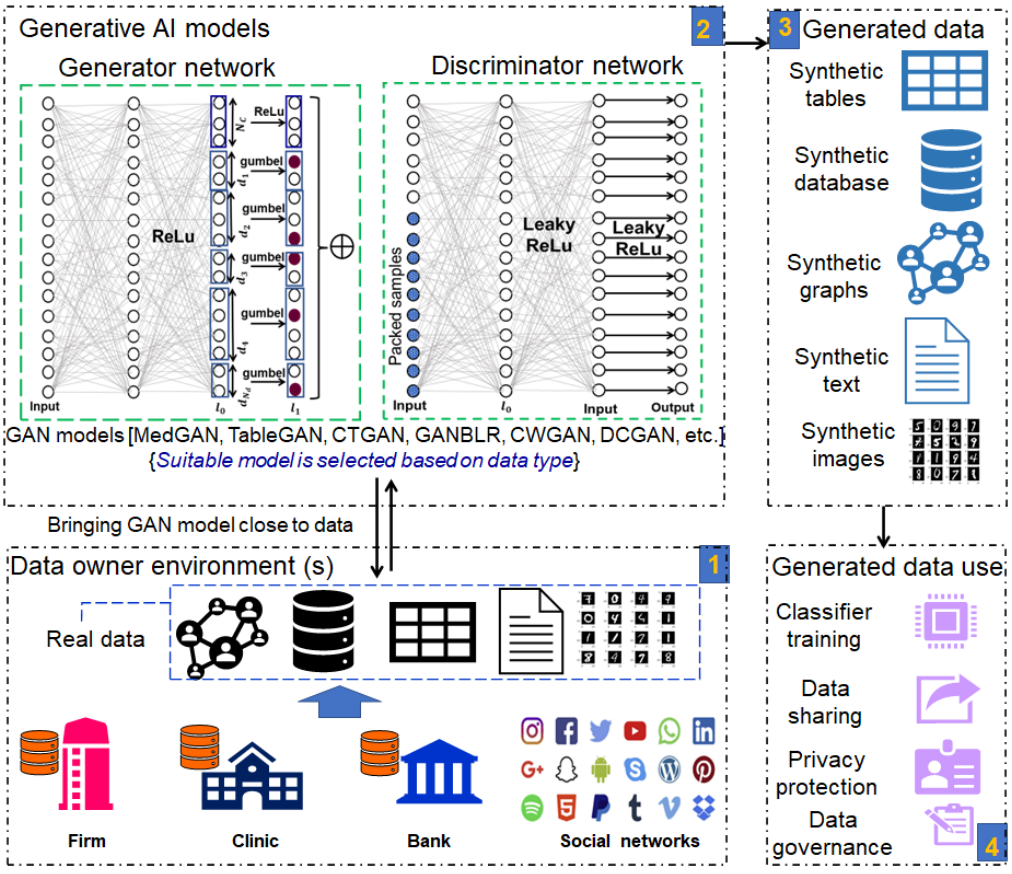


Figure 8. Overview of synthetic-data-based developments along with experimental details. In step 1, relevant data sources are identified and properties of real data are analyzed. In step 2, the appropriate GAN model is chosen and brought close to real data. In step 3, synthetic data is curated in the relevant modality. In step 4, the synthetic data is utilized either to train ML models or to share it for data mining.

5. Famous Privacy Enhancing Technologies That Are Widely Used for Privacy Preservation

Apart from SD and attribute-centric anonymization, there exist many privacy-enhancing technologies (PETs) that are used to secure personal data enclosed in diverse formats. Figure 9 presents notable PETs that are being used by data owners to secure personal data.

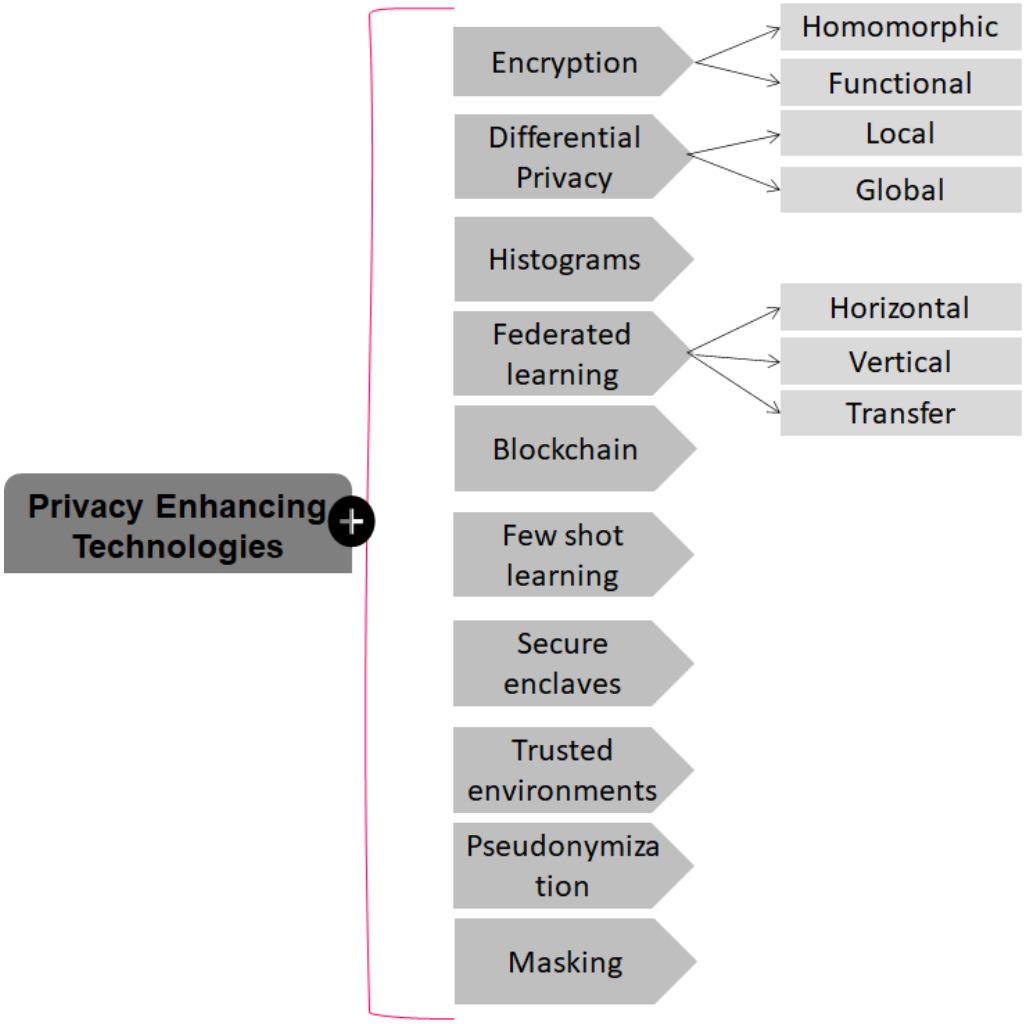


Figure 9. Overview of famous PETs used to secure personal data (e.g., privacy preservation).

Referring to Figure 9, encryption is widely used in dynamic environments (e.g., cloud computing) to prevent personal data from disclosure [122]. Recently, encryption techniques were amalgamated with AI techniques to preserve the privacy of personal data [123]. DP is widely used to preserve the privacy of data while permitting the analytics of data. It has become one of the famous PETs of modern times and is widely used in federated learning environments to preserve the privacy of data/parameters [124,125]. When privacy requirements are very high, only some statistics in the form of histograms are shared with the data analysts [126,127]. Federated learning is one of the modern PETs that does not aggregate personal data in central environments but allows AI model training. FL has become a SOTA solution for privacy preservation when data is located in various places [128,129]. Blockchain is another famous PET that can protect the privacy of the data as well as of participants. Blockchain is widely used to share data with multiple parties while preventing privacy breaches [130,131]. Few short learning is a famous PET for training large AI models with limited data, leading to privacy preservation in AI applications [132,133]. Trusted environment and secure enclaves are hardware-based approaches to guarantee the privacy of user data [134–136]. These approaches are becoming

famous PETs due to the rapid rise in AI-based applications. Pseudonymization is a widely used technique for privacy preservation in cloud-based environments such as smart home environments, smart grids, and medical domains [137,138]. Finally, masking techniques are widely used to hide identity or sensitive information in a variety of data formats (e.g., images, videos, etc.) [139,140]. All of these technologies play a vital role in preserving individual privacy in static or dynamic environments. In some cases, hybrid PET is used to provide stronger privacy guarantees than individual techniques [141]. Recently, many PETs were also developed to provide privacy guarantees in epidemic handling systems that are being developed to fight the ongoing pandemic [142–145]. Apart from these developments, more secure PETs are required to provide privacy guarantees in personal data handling.

6. Promising Future Research and Development Directions

In this section, we present promising research and development directions for future work. Figure 10 presents the list of promising directions that require more work from the research and development point of view. For example, there is a serious lack of methods that can be applied to poor-quality data. In some cases, the data sensed from some devices such as sensors may contain errors, and therefore, applying anonymity to them directly may lead to the wrong data mining results. To this end, some new methods were developed to add noise to true values only, leading to lowering errors in histogram generation, as well as generic data mining results [146]. Some works have also explored the impact of data quality on privacy in real-time applications and crowdsensing scenarios [147]. Recently, some data-centric developments have been made to augment the performance of AI applications involving fewer data [148]. Hence, it is vital to adopt data-centric practices in the privacy domain to develop more secure methods (data quality-aware anonymization) in the future era. In the coming years, AI models will be applied to every sector, and therefore, it is imperative to devise secure methods for privacy preservation in AI systems [149]. To this end, amalgamating anonymization methods with AI systems is an attractive area of research for the near future. Privacy and utility optimization is a long-standing problem in the privacy domain [150]. Hence, it is vital to devise methods that can provide strict privacy guarantees without compromising usefulness in data-sharing scenarios.

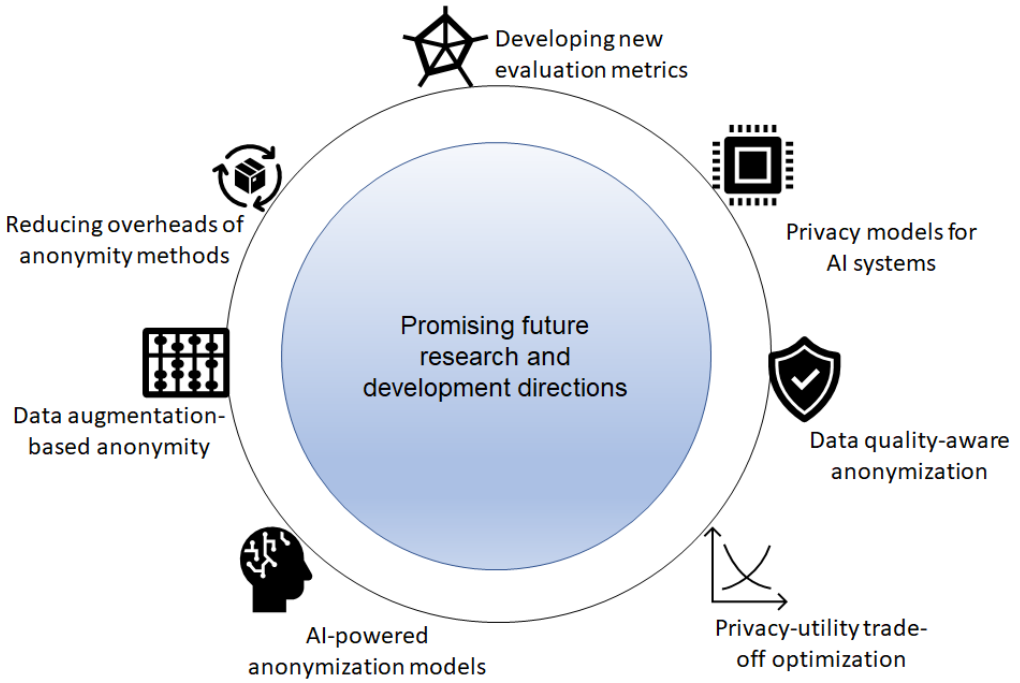


Figure 10. Promising direction for research or development in future endeavors.

Based on our recent work [151], we found that AI models can improve some critical parts of the conventional anonymization methods. Therefore, it is vital to amalgamate the AI methods with the traditional anonymization methods to secure personal data effectively. In the coming years, hybrid anonymization methods may be needed to preserve the privacy of data enclosed in multiple formats. In the past, most anonymization methods anonymized data without improving their structure (e.g., without balancing the distributions), leading to poor data quality and poor privacy guarantees. In the future, it is vital to improve/augment data before its anonymization, particularly when the quality of the data is poor [152]. Data augmentation-based anonymization methods can improve the quality of the data, leading to better data mining and analytical results. In most anonymization methods, the computing complexity rises with either the horizontal or vertical expansion of data, and therefore, it is extremely important to reduce computing overheads when the underlying data to be anonymized are large [153,154]. In future endeavors, devising low-cost methods (e.g., least computing and space complexity) for diverse data styles (e.g., graphs, images, text, etc.) is a vibrant area of research. Finally, there is a lack of metrics that can correctly capture the level of privacy and utility [155]. Hence, it is imperative to devise formally verified evaluation metrics in future endeavors. Finally, there is an increasing focus on developing clustering-based anonymization to improve the shortcomings of traditional anonymization methods [41,156]. However, there are various critical problems with these methods such as poor convergence, a substantial number of iterations, and higher computing overheads. Hence, improving the technical status, applicability, and convergence criteria is a vibrant area of research for the future. Figure 11 provides a detailed overview of the current challenges in the privacy domain and new research topics about preserving the privacy of personnel data. Due to the rapid rise in personal data, privacy issues are becoming more evident [157,158]. Therefore, robust, efficient, and low-cost anonymization methods are necessary to preserve privacy in futuristic applications. The analysis discussed in Figure 11 provides a clear overview of the future research trajectories in preserving the privacy of personal data.

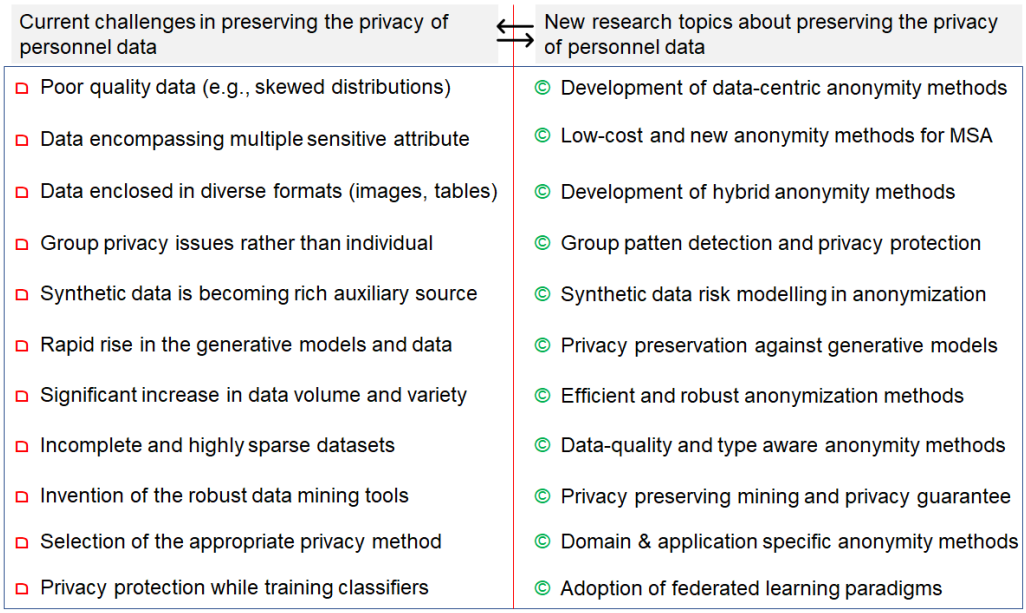


Figure 11. Detailed overview of the current challenges in the privacy domain and new research topics about preserving the privacy of personnel data.

It is worth noting that most of the existing privacy preservation methods have focused on preserving three key properties (e.g., identity information, sensitive information, and membership information) in personal data handling. However, in some cases, an individual can have multiple types of sensitive information in their data, and therefore, there is a

need for privacy preservation methods that can provide privacy protection for multiple types of sensitive information. In addition, due to the extensive use of digital technologies, the privacy of different sensitive readings (i.e., blood pressure, heart rate) also requires protection from prying eyes [159,160]. Hence, it is vital to devise more secure methods that can preserve the properties of the personal data encompassed in different formats, as well as stemming from diverse domains.

7. Conclusions and Future Work

This paper presented a painstaking analysis of the latest and state-of-the-art (SOTA) developments in the information privacy domain. Specifically, we identified and described multiple research tracks for the information privacy domain with a special focus on SOTA attribute-centric anonymization methods that were recently developed to balance privacy and utility. Later, we described SOTA synthetic data generation methods that were rigorously used to meet the privacy and data analytics demands in the modern era. To the best of our knowledge, both these categories (i.e., attribute-centric anonymization methods and synthetic data generation methods) of privacy-enhancing technologies have not been thoroughly covered in the current literature. Furthermore, we demonstrate famous privacy solutions other than these two methods that are widely used to secure personal data encompassed in multiple formats. The promising avenues for future research were also discussed to foster further developments/research in the privacy preservation area. The contents enclosed in this paper can pave the way for future development in the privacy domain. Finally, this work aligns with the recent trends toward responsible data science (e.g., fair, transparent, and privacy-preserved use of personal data) and can open up avenues for future research and development. This paper encloses much-needed knowledge concerning information privacy and can provide a strong base for future work in the information privacy domain. In the future, we intend to cover generative methods for other data styles (e.g., images, text, audio, etc.) and data augmentation strategies (e.g., random, selective, sampling, etc.) used in improving the quality of the bad data. Finally, we aim to pinpoint existing developments concerning privacy-preserving synthetic data generation using differential privacy-based techniques.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

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

Acknowledgments: The authors thank the expert reviewers who thoroughly evaluated this article and provided very constructive feedback, which significantly enhanced the quality of this article.

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

References

1. Wieringa, J.; Kannan, P.K.; Ma, X.; Reutterer, T.; Risselada, H.; Skiera, B. Data analytics in a privacy-concerned world. *J. Bus. Res.* **2021**, *122*, 915–925. [CrossRef]
2. Sweeney, L. Simple demographics often identify people uniquely. *Health* **2000**, *671*, 1–34.
3. Majeed, A.; Lee, S. Anonymization techniques for privacy preserving data publishing: A comprehensive survey. *IEEE Access* **2020**, *9*, 8512–8545. [CrossRef]
4. Sweeney, L. k-anonymity: A model for protecting privacy. *Int. J. Uncertain. Fuzziness-Knowl.-Based Syst.* **2002**, *10*, 557–570. [CrossRef]
5. Machanavajjhala, A.; Kifer, D.; Gehrke, J.; Venkitasubramaniam, M. l-diversity: Privacy beyond k-anonymity. *ACM Trans. Knowl. Discov. Data (TKDD)* **2007**, *1*, 3-es. [CrossRef]
6. Li, N.; 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; pp. 106–115.
7. Dwork, C. Differential privacy: A survey of results. In *International Conference on Theory and Applications of Models of Computation*; Springer: Berlin/Heidelberg, Germany, 2008; pp. 1–19.

8. Chen, Y.; Gan, W.; Wu, Y.; Philip, S.Y. Privacy-Preserving Federated Mining of Frequent Itemsets. *Inf. Sci.* **2023**, *625*, 504–520. [CrossRef]
9. Qiu, S.; Pi, D.; Wang, Y.; Liu, Y. Novel trajectory privacy protection method against prediction attacks. *Expert Syst. Appl.* **2023**, *213*, 118870. [CrossRef]
10. Kaur, H.; Hooda, N.; Singh, H. *k*-anonymization of social network data using Neural Network and SVM: K-NeuroSVM. *J. Inf. Secur. Appl.* **2023**, *72*, 103382. [CrossRef]
11. Payton, T.; Claypoole, T. *Privacy in the Age of Big Data: Recognizing Threats, Defending Your Rights, and Protecting Your Family*; Rowman & Littlefield: Lanham, MD, USA, 2023.
12. Majeed, A.; Hwang, S.O. When AI meets Information Privacy: The Adversarial Role of AI in Data Sharing Scenario. *IEEE Access* **2023**, *11*, 76177–76195. [CrossRef]
13. Song, J.; Wang, W.; Gadekallu, T.R.; Cao, J.; Liu, Y. Eppda: An efficient privacy-preserving data aggregation federated learning scheme. *IEEE Trans. Netw. Sci. Eng.* **2022**. https: //doi.org/10.1109/TNSE.2022.3153519. [CrossRef]
14. Strickland, E. Andrew Ng, AI Minimalist: The Machine-Learning Pioneer Says Small is the New Big. *IEEE Spectrum.* **2022**, *59*, 22–50. [CrossRef]
15. Whang, S.E.; Roh, Y.; Song, H.; Lee, J.G. Data collection and quality challenges in deep learning: A data-centric ai perspective. *VLDB J.* **2023**, *32*, 791–813. [CrossRef]
16. El Emam, K. Seven ways to evaluate the utility of synthetic data. *IEEE Secur. Priv.* **2020**, *18*, 56–59. [CrossRef]
17. Li, J.; Cairns, B.J.; Li, J.; Zhu, T. Generating synthetic mixed-type longitudinal electronic health records for artificial intelligent applications. *Npj Digit. Med.* **2023**, *6*, 98. [CrossRef] [PubMed]
18. Figueira, A.; Vaz, B. Survey on synthetic data generation, evaluation methods and GANs. *Mathematics* **2022**, *10*, 2733. [CrossRef]
19. James, S.; Harbron, C.; Branson, J.; Sundler, M. Synthetic data use: Exploring use cases to optimise data utility. *Discov. Artif. Intell.* **2021**, *1*, 15. [CrossRef]
20. Hoang, A.T.; Carminati, B.; Ferrari, E. Protecting Privacy in Knowledge Graphs with Personalized Anonymization. *IEEE Trans. Dependable Secur. Comput.* **2023**. [CrossRef]
21. Fan, Y.; Shi, X.; Zhang, S.; Tong, Y. Anonymous Methods Based on Multi-Attribute Clustering and Generalization Constraints. *Electronics* **2023**, *12*, 1897. [CrossRef]
22. Yao, L.; Wang, X.; Hu, H.; Wu, G. A Utility-aware Anonymization Model for Multiple Sensitive Attributes Based on Association Concealment. *IEEE Trans. Dependable Secur. Comput.* **2023**. [CrossRef]
23. De Pascale, D.; Cascavilla, G.; Tamburri, D.A.; Van Den Heuvel, W.J. Real-world K-Anonymity applications: The KGen approach and its evaluation in fraudulent transactions. *Inf. Syst.* **2023**, *115*, 102193. [CrossRef]
24. Aldeen, Y.A.A.S.; Salleh, M.; Razzaque, M.A. A comprehensive review on privacy preserving data mining. *SpringerPlus* **2015**, *4*, 1–36. [CrossRef]
25. Mendes, R.; Vilela, J.P. Privacy-preserving data mining: Methods, metrics, and applications. *IEEE Access* **2017**, *5*, 10562–10582. [CrossRef] [PubMed]
26. Rathi, M.; Rajavat, A. Analysing Cryptographic and Random Data Sanitization Techniques in Privacy Preserving Data Mining. In *Emerging Strategies in Research—Going Beyond Disciplinary Boundaries*; Allied Publishers: New Delhi, India, 2023; Volume 83. [CrossRef]
27. Naresh, V.S.; Thamarai, M. Privacy-preserving data mining and machine learning in healthcare: Applications, challenges, and solutions. *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.* **2023**, *13*, e1490.
28. Gao, Y.; Chen, L.; Han, J.; Wu, G.; Susilo, W. IoT Privacy-preserving Data Mining with Dynamic Incentive Mechanism. *IEEE Internet Things J.* **2023**. [CrossRef]
29. Feng, J.; Yang, L.T.; Ren, B.; Zou, D.; Dong, M.; Zhang, S. Tensor recurrent neural network with differential privacy. *IEEE Trans. Comput.* **2023**. [CrossRef]
30. Karanasios, S. The pursuit of relevance and impact: A review of the immediate response of the information systems field to COVID-19. *Inf. Syst. J.* **2022**, *32*, 856–887. [CrossRef]
31. Antons, D.; Breidbach, C.F.; Joshi, A.M.; Salge, T.O. Computational literature reviews: Method, algorithms, and roadmap. *Organ. Res. Methods* **2023**, *26*, 107–138. [CrossRef]
32. Carvalho, T.; Moniz, N.; Faria, P.; Antunes, L. Survey on Privacy-Preserving Techniques for Data Publishing. *arXiv* **2022**, arXiv:2201.08120. [CrossRef]
33. Pujol, D.; Machanavajjhala, A. Equity and Privacy: More Than Just a Tradeoff. *IEEE Secur. Priv.* **2021**, *19*, 93–97.
34. Cao, X.; Cao, Y.; Pappachan, P.; Nakamura, A.; Yoshikawa, M. Differentially Private Streaming Data Release Under Temporal Correlations via Post-processing. In *IFIP Annual Conference on Data and Applications Security and Privacy*; Springer Nature: Cham, Switzerland, 2023; pp. 184–200. [CrossRef]
35. Torra, V.; Navarro-Arribas, G. Attribute disclosure risk for k-anonymity: The case of numerical data. *Int. J. Inf. Secur.* **2023**, 1–10. [CrossRef]
36. Hathaliya, J.J.; Tanwar, S. An exhaustive survey on security and privacy issues in Healthcare 4.0. *Comput. Commun.* **2020**, *153*, 311–335. [CrossRef]
37. Srinivasan, K.; Rathee, G.; Raja, M.R.; Jaglan, N.; Mahendiran, T.V.; Palaniswamy, T. Secure multimedia data processing scheme in medical applications. *Multimed. Tools Appl.* **2022**, *81*, 9079–9090. [CrossRef]

38. Liu, B.; Ding, M.; Shaham, S.; Rahayu, W.; Farokhi, F.; Lin, Z. When machine learning meets privacy: A survey and outlook. *Acm Comput. Surv. (CSUR)* **2021**, *54*, 1–36. [CrossRef]
39. Gadad, V.; Sowmyarani, C.N. Incremental Diversity: An Efficient Anonymization Technique for PPDP of Multiple Sensitive Attributes. *Int. J. Adv. Comput. Sci. Appl.* **2023**, *14*, 3. [CrossRef]
40. Stadler, T.; Oprisanu, B.; Troncoso, C. Synthetic data–anonymisation groundhog day. In Proceedings of the 31st USENIX Security Symposium (USENIX Security 22), BOSTON, MA, USA, 10–12 August 2022; pp. 1451–1468. [CrossRef]
41. Chen, L.; Zeng, L.; Mu, Y.; Chen, L. Global Combination and Clustering based Differential Privacy Mixed Data Publishing. *IEEE Trans. Knowl. Data Eng.* **2023**. https://doi.org/10.1109/TKDE.2023.3237822.
42. Chakraborty, C.; Othman, S.B.; Almalki, F.A.; Sakli, H. FC-SEEDA: Fog computing-based secure and energy efficient data aggregation scheme for Internet of healthcare Things. *Neural Comput. Appl.* **2023**, 1–17. [CrossRef]
43. Li, W.; Ding, W.; Sadasivam, R.; Cui, X.; Chen, P. His-GAN: A histogram-based GAN model to improve data generation quality. *Neural Netw.* **2019**, *119*, 31–45. [CrossRef]
44. Liu, H.; Tian, Y.; Peng, C.; Wu, Z. Privacy-utility equilibrium data generation based on Wasserstein generative adversarial networks. *Inf. Sci.* **2023**, *642*, 119069. [CrossRef]
45. Ren, Z.; Zhu, Y.; Liu, Z.; Feng, K. Few-shot GAN: Improving the performance of intelligent fault diagnosis in severe data imbalance. *IEEE Trans. Instrum. Meas.* **2023**, *72*, 3516814. [CrossRef]
46. Hongbin, F.; Zhi, Z. Privacy-Preserving Data Aggregation Scheme Based on Federated Learning for IIoT. *Mathematics* **2023**, *11*, 214. [CrossRef]
47. Paul, M.; Maglaras, L.; Ferrag, M.A.; AlMomani, I. Digitization of healthcare sector: A study on privacy and security concerns. *ICT Express* **2023**, *9*, 571–588. [CrossRef]
48. Muneeswari, G.; Varun, S.S.; Hegde, R.; Priya, S.S.; Shermila, P.J.; Prasanth, A. Self-diagnosis platform via IOT-based privacy preserving medical data. *Meas. Sens.* **2023**, *25*, 100636. [CrossRef]
49. Xie, H.; Zheng, J.; He, T.; Wei, S.; Hu, C. TEBDS: A Trusted Execution Environment-and-Blockchain-supported IoT data sharing system. *Future Gener. Comput. Syst.* **2023**, *140*, 321–330. [CrossRef]
50. Liu, C.; Chen, S.; Zhou, S.; Guan, J.; Ma, Y. A novel privacy preserving method for data publication. *Inf. Sci.* **2019**, *501*, 421–435. [CrossRef]
51. Hewage, U.H.W.A.; Sinha, R.; Naeem, M.A. Privacy-preserving data (stream) mining techniques and their impact on data mining accuracy: A systematic literature review. *Artif. Intell. Rev.* **2023**, *56*, 10427–10464. [CrossRef]
52. Terziyan, V.; Malyk, D.; Golovianko, M.; Branytskyi, V. Encryption and Generation of Images for Privacy-Preserving Machine Learning in Smart Manufacturing. *Procedia Comput. Sci.* **2023**, *217*, 91–101. [CrossRef]
53. Qin, C.; Wu, L.; Meng, W.; Xu, Z.; Li, S.; Wang, H. A privacy-preserving blockchain-based tracing model for virus-infected people in cloud. *Expert Syst. Appl.* **2023**, *211*, 118545. [CrossRef]
54. Kumuthini, J.; Zass, L.; Chaouch, M.; Fadlelmola, F.M.; Mulder, N.; Radouani, F.; Ras, V.; Samtal, C.; Tchamga, M.S.; Sathan, D.; et al. *Genomics Data Sharing*; Academic Press: Cambridge, MA, USA, 2023; pp. 111–135. [CrossRef]
55. Yang, X.; Yi, X.; Kelarev, A.; Rylands, L.; Lin, Y.; Ryan, J. Protecting Private Information for Two Classes of Aggregated Database Queries. *Informatics* **2022**, *9*, 66.
56. Lin, C.-H.; Yu, C.-M.; Huang, C.-Y. DPView: Differentially Private Data Synthesis Through Domain Size Information. *IEEE Internet Things J.* **2022**, *9*, 15886–15900. [CrossRef]
57. Sei, Y.; Andrew, J.; Okumura, H.; Ohsuga, A. Privacy-preserving collaborative data collection and analysis with many missing values. *IEEE Trans. Dependable Secur. Comput.* **2023**, *20*, 2158–2173. [CrossRef]
58. Krishna, S.; Murthy, U.V. Evolutionary tree-based quasi identifier and federated gradient privacy preservations over big healthcare data. *Int. J. Electr. Comput. Eng.* **2022**, *12*, 903. [CrossRef]
59. Chong, K.M.; Malip, A. Bridging unlinkability and data utility: Privacy preserving data publication schemes for healthcare informatics. *Comput. Commun.* **2022**, *191*, 194–207. [CrossRef]
60. Breger, A.; Selby, I.; Roberts, M.; Babar, J.; Gkrania-Klotsas, E.; Preller, J.; Sánchez, L.E.; Rudd, J.H.F.; Aston, J.A.D.; Weir-McCall, J.R. A pipeline to further enhance quality, integrity and reusability of the NCCID clinical data. *Sci. Data* **2023**, *10*, 493. [CrossRef]
61. Majeed, A. Attribute-centric anonymization scheme for improving user privacy and utility of publishing e-health data. *J. King Saud-Univ.-Comput. Inf. Sci.* **2019**, *31*, 426–435. [CrossRef] [PubMed]
62. Onesimu, J.A.; Karthikeyan, J.; Eunice, J.; Pomplun, M.; Dang, H. Privacy preserving attribute-focused anonymization scheme for healthcare data publishing. *IEEE Access* **2022**, *10*, 86979–86997. [CrossRef]
63. Hui, T.; Wee-Chung, L.A.; Earnest, F. A scheme of hybrid privacy protection and utility levels for medical data. *arXiv* **2022**, arXiv:2204.13880. [CrossRef]
64. Khan, M.S.; Anjum, A.; Saba, T.; Rehman, A.; Tariq, U. Improved generalization for secure personal data publishing using deviation. *IT Prof.* **2021**, *23*, 75–80.
65. Zhang, S.; Li, X. Differential privacy medical data publishing method based on attribute correlation. *Sci. Rep.* **2022**, *12*, 15725. [CrossRef]
66. Sadhya, D.; Chakraborty, B. Quantifying the Effects of Anonymization Techniques over Micro-databases. *IEEE Trans. Emerg. Top. Comput.* **2022**, *10*, 1979–1992. [CrossRef]

67. Srijayanthi, S.; Sethukarasi, T. Design of privacy preserving model based on clustering involved anonymization along with feature selection. *Comput. Secur.* **2023**, *126*, 103027. [CrossRef]
68. Canbay, Y.; Sagiroglu, S.; Vural, Y. A Mondrian-based Utility Optimization Model for Anonymization. In Proceedings of the 2019 4th International Conference on Computer Science and Engineering (UBMK), Samsun, Turkey, 11–15 September 2019; pp. 709–714. [CrossRef]
69. Arca, S.; Hewett, R. Analytics on anonymity for privacy retention in smart health data. *Future Internet* **2021**, *13*, 274.
70. Zouinina, S.; Bennani, Y.; Rogovschi, N.; Lyhyaoui, A. Data anonymization through collaborative multi-view microaggregation. *J. Intell. Syst.* **2020**, *30*, 327–345. [CrossRef]
71. Yan, Y.; Eyeleko, A.H.; Mahmood, A.; Li, J.; Dong, Z.; Xu, F. Privacy preserving dynamic data release against synonymous linkage based on microaggregation. *Sci. Rep.* **2022**, *12*, 2352. [CrossRef]
72. Lee, S.; Shin, W.Y. Utility-Embraced Microaggregation for Machine Learning Applications. *IEEE Access* **2022**, *10*, 64535–64546. [CrossRef] [PubMed]
73. Yan, Y.; Sun, Z.; Mahmood, A.; Xu, F.; Dong, Z.; Sheng, Q.Z. Achieving Differential Privacy Publishing of Location-Based Statistical Data Using Grid Clustering. *ISPRS Int. J. Geo-Inf.* **2022**, *11*, 404. [CrossRef]
74. Thanga Revathi, S.; Gayathri, A.; Kalaivani, J.; Christo, M.S.; Pelusi, D.; Azees, M. Cloud-Assisted Privacy-Preserving Method for Healthcare Using Adaptive Fractional Brain Storm Integrated Whale Optimization Algorithm. *Secur. Commun. Netw.* **2021**, *2021*, 6210054. [CrossRef]
75. Anand, K.; Vijayaraj, A.; Vijay Anand, M. Privacy preserving framework using Gaussian mutation based firebug optimization in cloud computing. *J. Supercomput.* **2022**, *1*, 1–24. [CrossRef]
76. Anand, K.; Vijayaraj, A.; Vijay Anand, M. An enhanced bacterial foraging optimization algorithm for secure data storage and privacy-preserving in cloud. *Peer-to-Peer Netw. Appl.* **2022**, *15*, 2007–2020.
77. Canbay, Y.; Vural, Y.; Sagiroglu, Ş. OAN: Outlier record-oriented utility-based privacy preserving model. *J. Fac. Eng. Archit. Gazi Univ.* **2020**, *35*, 355–368. [CrossRef]
78. Onesimu, J.A.; 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.
79. Xie, M.; Huang, M.; Bai, Y.; Hu, Z. The anonymization protection algorithm based on fuzzy clustering for the ego of data in the internet of things. *J. Electr. Comput. Eng.* **2017**, *2017*, 2970673. [CrossRef]
80. Yang, L.; 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. [CrossRef]
81. Li, H.; Guo, F.; Zhang, W.; Wang, J.; Xing, J. (a, k)-Anonymous scheme for privacy-preserving data collection in IoT-based healthcare services systems. *J. Med. Syst.* **2018**, *42*, 1–9. [CrossRef]
82. Altameem, A.; Kovtun, V.; Al-Ma'aitah, M.; Altameem, T.; Fouad, H.; Youssef, A.E. Patient's data privacy protection in medical healthcare transmission services using back propagation learning. *Comput. Electr. Eng.* **2022**, *102*, 108087. [CrossRef]
83. Nayahi, J.J.; Kavitha, V. Privacy and utility preserving data clustering for data anonymization and distribution on Hadoop. *Future Gener. Comput. Syst.* **2017**, *74*, 393–408. [CrossRef]
84. Ashkouti, F.; 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. [CrossRef]
85. Solanki, P.; Garg, S.; Chhikaniwala, H. Preserve Privacy on Streaming Data During the Process of Mining Using User Defined Delta Value. In *Innovative Data Communication Technologies and Application: Proceedings of ICIDCA 2021, Singapore, 24 February 2022*; Springer Nature Singapore: Singapore, 2022; pp. 197–212. [CrossRef]
86. 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.
87. Liu, F.; Cheng, Z.; Chen, H.; Wei, Y.; Nie, L.; Kankanhalli, M. Privacy-preserving synthetic data generation for recommendation systems. In Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval, Madrid, Spain, 6 July 2022; pp. 1379–1389.
88. Rankin, D.; Black, M.; Bond, R.; Wallace, J.; Mulvenna, M.; Epelde, G. Reliability of supervised machine learning using synthetic data in health care: Model to preserve privacy for data sharing. *JMIR Med. Inform.* **2020**, *8*, e18910.
89. Chen, R.J.; Lu, M.Y.; Chen, T.Y.; Williamson, D.F.; Mahmood, F. Synthetic data in machine learning for medicine and healthcare. *Nat. Biomed. Eng.* **2021**, *5*, 493–497. [CrossRef]
90. Hahn, W.; Schütte, K.; Schultz, K.; Wolkenhauer, O.; Sedlmayr, M.; Schuler, U.; Eichler, M.; Bej, S.; Wolfien, M. Contribution of Synthetic Data Generation towards an Improved Patient Stratification in Palliative Care. *J. Pers. Med.* **2022**, *12*, 1278. [CrossRef]
91. Gonzales, A.; Guruswamy, G.; Smith, S.R. Synthetic data in health care: A narrative review. *PLoS Digit. Health* **2023**, *2*, e0000082. [CrossRef]
92. Chen, X.; Wang, C.; Yang, Q.; Hu, T.; Jiang, C. Locally differentially private high-dimensional data synthesis. *Sci. China Inf. Sci.* **2023**, *66*, 1–8. [CrossRef] [PubMed]
93. De Cristofaro, E. What Is Synthetic Data? The Good, The Bad, and The Ugly. *arXiv* **2023**, arXiv:2303.01230. [CrossRef]
94. Harder, F.; Adamczewski, K.; Park, M. Dp-merf: Differentially private mean embeddings with randomfeatures for practical privacy-preserving data generation. In Proceedings of the International Conference on Artificial Intelligence and Statistics, San Diego, CA, USA, 18 March 2021; pp. 1819–1827.

95. Lee, J.; Kim, M.; Jeong, Y.; Ro, Y. Differentially Private Normalizing Flows for Synthetic Tabular Data Generation. In Proceedings of the AAAI Conference on Artificial Intelligence, Palo Alto, CA, USA, 28 June 2022; Volume 36, pp. 7345–7353.
96. Jälkö, J.; Lagerspetz, E.; Haukka, J.; Tarkoma, S.; Honkela, A.; Kaski, S. Privacy-preserving data sharing via probabilistic modeling. *Patterns* **2021**, *2*, 100271.
97. Yale, A.; Dash, S.; Dutta, R.; Guyon, I.; Pavao, A.; Bennett, K.P. Generation and evaluation of privacy preserving synthetic health data. *Neurocomputing* **2020**, *416*, 244–255. [CrossRef]
98. Lenatti, M.; Paglialonga, A.; Orani, V.; Ferretti, M.; Mongelli, M. Characterization of Synthetic Health Data Using Rule-Based Artificial Intelligence Models. *IEEE J. Biomed. Health Inform.* **2023**, *27*, 3760–3769. [CrossRef]
99. Bhanot, K.; Pedersen, J.; Guyon, I.; Bennett, K.P. Investigating synthetic medical time-series resemblance. *Neurocomputing* **2022**, *494*, 368–378. [CrossRef]
100. Nikolentzos, G.; Vazirgiannis, M.; Xypolopoulos, C.; Lingman, M.; Brandt, E. Synthetic electronic health records generated with variational graph autoencoders. *NPJ Digit. Med.* **2023**, *6*, 83. [CrossRef]
101. Hernandez, M.; Epelde, G.; Beristain, A.; Álvarez, R.; Molina, C.; Larrea, X.; Alberdi, A.; Timoleon, M.; Bamidis, P.; Konstantinidis, E. Incorporation of synthetic data generation techniques within a controlled data processing workflow in the health and wellbeing domain. *Electronics* **2022**, *11*, 812. [CrossRef]
102. Kuo, N.I.; Polizzotto, M.N.; Finfer, S.; Garcia, F.; Sönnerborg, A.; Zazzi, M.; Böhm, M.; Kaiser, R.; Jorm, L.; Barbieri, S. The Health Gym: Synthetic health-related datasets for the development of reinforcement learning algorithms. *Sci. Data* **2022**, *9*, 693. [CrossRef]
103. Rodriguez-Almeida, A.J.; Fabelo, H.; Ortega, S.; Deniz, A.; Balea-Fernandez, F.J.; Quevedo, E.; Soguero-Ruiz, C.; Wägner, A.M.; Callico, G.M. Synthetic Patient Data Generation and Evaluation in Disease Prediction Using Small and Imbalanced Datasets. *IEEE J. Biomed. Health Inform.* **2022**, *27*, 2670–2680. [CrossRef]
104. Banerjee, S.; Bishop, T.R. dsSynthetic: Synthetic data generation for the DataSHIELD federated analysis system. *BMC Res. Notes* **2022**, *15*, 230. [CrossRef]
105. Larrea, X.; Hernandez, M.; Epelde, G.; Beristain, A.; Molina, C.; Alberdi, A.; Rankin, D.; Bamidis, P.; Konstantinidis, E. Synthetic Subject Generation with Coupled Coherent Time Series Data. *Eng. Proc.* **2022**, *18*, 7. [CrossRef] [PubMed]
106. Venugopal, R.; Shafqat, N.; Venugopal, I.; Tillbury, B.M.; Stafford, H.D.; Bourazeri, A. Privacy preserving Generative Adversarial Networks to model Electronic Health Records. *Neural Netw.* **2022**, *153*, 339–348.
107. Yoshikawa, H.; Uchiyama, A. Privacy-preserving data augmentation for thermal sensation dataset based on variational autoencoder. In Proceedings of the 9th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation, Boston, MA, USA, 9 November 2022; pp. 286–287. [CrossRef]
108. Hyun, J.; Lee, Y.; Son, H.M.; Lee, S.H.; Pham, V.; Park, J.U.; Chung, T.M. Synthetic Data Generation System for AI-Based Diabetic Foot Diagnosis. *SN Comput. Sci.* **2021**, *2*, 345.
109. Zhang, A.; Xing, L.; Zou, J.; Wu, J.C. Shifting machine learning for healthcare from development to deployment and from models to data. *Nat. Biomed. Eng.* **2022**, *4*, 1–6. [CrossRef]
110. Yale, A.; Dash, S.; Bhanot, K.; Guyon, I.; Erickson, J.S.; Bennett, K.P. Synthesizing quality open data assets from Private Health Research Studies. In *Business Information Systems Workshops: BIS 2020 International Workshops*, Colorado Springs, CO, USA, 8–10 June 2020; pp. 324–335. [CrossRef] [PubMed]
111. Narteni, S.; Orani, V.; Ferrari, E.; Verda, D.; Cambiaso, E.; Mongelli, M. A New XAI-based Evaluation of Generative Adversarial Networks for IMU Data Augmentation. In Proceedings of the 2022 IEEE International Conference on E-health Networking, Application & Services (HealthCom), Genoa, Italy, 17–19 October 2022; pp. 167–172.
112. Raveendran, R.; Raj, E.D. Deep Generative Models Under GAN: Variants, Applications, and Privacy Issues. In Proceedings of the 7th International Conference on Information System Design and Intelligent Applications (India 2022), Hyderabad, India, 25–26 February 2022; Volume 28, pp. 93–105.
113. Li, M.; Zhuang, D.; Chang, J.M. MC-GEN: Multi-level clustering for private synthetic data generation. *Knowl.-Based Syst.* **2023**, *21*, 110239.
114. Shahani, S.; Abraham, J. Techniques for Privacy-Preserving Data Aggregation in an Untrusted Distributed Environment. In Proceedings of the 6th Joint International Conference on Data Science & Management of Data (10th ACM IKDD CODS and 28th COMAD), Mumbai, India, 4–7 January 2023; pp. 286–287. [CrossRef]
115. Hindistan, Y.S.; Yetkin, E.F. A Hybrid Approach with GAN and DP for Privacy Preservation of IIoT Data. *IEEE Access*. **2023**, *1*, 5837–5849.
116. Zhang, Y.; Zhu, Y.; Zhou, Y.; Yuan, J. Frequency Estimation Mechanisms under (ϵ, δ)-Utility-optimized Local Differential Privacy. *IEEE Trans. Emerg. Top. Comput.* **2023**. [CrossRef]
117. Nair, A.K.; Sahoo, J.; Raj, E.D. Privacy preserving Federated Learning framework for IoMT based big data analysis using edge computing. *Comput. Stand. Interfaces* **2023**, *4*, 103720. [CrossRef]
118. Zhang, Z.; Xu, X.; Xiao, F. LGAN-DP: A novel differential private publication mechanism of trajectory data. *Future Gener. Comput. Syst.* **2023**, *141*, 692–703. [CrossRef]
119. Duan, S.; Liu, C.; Han, P.; Jin, X.; Zhang, X.; He, T.; Pan, H.; Xiang, X. HT-Fed-GAN: Federated Generative Model for Decentralized Tabular Data Synthesis. *Entropy* **2023**, *25*, 88. [CrossRef]

120. Cheng, K.; Tahir, R.; Eric, L.K.; Li, M. An analysis of generative adversarial networks and variants for image synthesis on MNIST dataset. *Multimed. Tools Appl.* **2020**, *79*, 13725–13752. [CrossRef]
121. Castelli, M.; Manzoni, L. Generative models in artificial intelligence and their applications. *Appl. Sci.* **2022**, *12*, 4127. [CrossRef]
122. Ma, Z.; Wang, J.; Gai, K.; Duan, P.; Zhang, Y.; Luo, S. Fully homomorphic encryption-based privacy-preserving scheme for cross edge blockchain network. *J. Syst. Archit.* **2023**, *134*, 102782. [CrossRef]
123. Zhang, M.; Huang, S.; Shen, G.; Wang, Y. PPNNP: A privacy-preserving neural network prediction with separated data providers using multi-client inner-product encryption. *Comput. Stand. Interfaces* **2023**, *84*, 103678. [CrossRef]
124. Li, X.; Xu, L.; Zhang, H.; Xu, Q. Differential privacy preservation for graph auto-encoders: A novel anonymous graph publishing model. *Neurocomputing* **2023**, *521*, 113–125. [CrossRef]
125. Guo, S.; Wang, X.; Long, S.; Liu, H.; Hai, L.; Sam, T.H. A federated learning scheme meets dynamic differential privacy. *CAAI Trans. Intell. Technol.* **2023**. [CrossRef]
126. Liu, X.; Chen, Y. Group effect-based privacy-preserving data aggregation for mobile crowdsensing. *Comput. Netw.* **2023**, *222*, 109507. [CrossRef]
127. Chen, Q.; Ni, Z.; Zhu, X.; Xia, P. Differential privacy histogram publishing method based on dynamic sliding window. *Front. Comput. Sci.* **2023**, *17*, 174809. [CrossRef]
128. Gao, H.; He, N.; Gao, T. SVeriFL: Successive verifiable federated learning with privacy-preserving. *Inf. Sci.* **2023**, *622*, 98–114. [CrossRef]
129. Ouyang, L.; Wang, F.-Y.; Tian,Y.; Jia, X.; Qi, H.; Wang, G. Artificial identification: A novel privacy framework for federated learning based on blockchain. *IEEE Trans. Comput. Soc. Syst.* **2023**. [CrossRef]
130. Singh, S.K.; Yang, L.T.; Park, J.H. FusionFedBlock: Fusion of blockchain and federated learning to preserve privacy in industry 5.0. *Inf. Fusion* **2023**, *90*, 233–240. [CrossRef]
131. Liu, M.; Zhang, Z.; Chai, W.; Wang, B. Privacy-preserving COVID-19 contact tracing solution based on blockchain. *Comput. Stand. Interfaces* **2023**, *83*, 103643. [CrossRef]
132. Raveendran, A.; Dhanapal, R. A non-interactive privacy preserved training technique based on hybrid deep learning. *Optik* **2023**, *273*, 170420. [CrossRef] [PubMed]
133. Cai, H.; Zhu, X.; Wen, P.; Han, W.; Wu, L. A Survey of Few-Shot Learning for Image Classification of Aerial Objects. In *China Aeronautical Science and Technology Youth Science Forum*; Springer Nature: Singapore, 2023; pp. 570–582. [CrossRef]
134. Zhang, D.; Ren, L.; Shafiq, M.; Gu, Z. A Privacy Protection Framework for Medical Image Security without Key Dependency Based on Visual Cryptography and Trusted Computing. *Comput. Intell. Neurosci.* **2023**, *2023*, 6758406.
135. Huang, P.H.; Tu, C.H.; Chung, S.M.; Wu, P.Y.; Tsai, T.L.; Lin, Y.A.; Dai, C.Y.; Liao, T.Y. SecureTVM: A TVM-Based Compiler Framework for Selective Privacy-Preserving Neural Inference. *ACM Trans. Des. Autom. Electron. Syst.* **2023**, *28*, 1–28. [CrossRef]
136. Chen, K. Confidential High-Performance Computing in the Public Cloud. *IEEE Internet Comput.* **2023**, *27*, 24–32. [CrossRef]
137. Abdul-Jabbar, M.D.; Aldeen, Y.A. State-of-the-Art in Data Integrity and Privacy-Preserving in Cloud Computing. *J. Eng.* **2023**, *29*, 42–60. [CrossRef]
138. Tall, A.M.; Zou, C.C. A Framework for Attribute-Based Access Control in Processing Big Data with Multiple Sensitivities. *Appl. Sci.* **2023**, *13*, 1183. [CrossRef]
139. Kunchala, A.; Bouroche, M.; Schoen-Phelan, B. Towards A Framework for Privacy-Preserving Pedestrian Analysis. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, Waikoloa, HI, USA, 2–7 January 2023; pp. 4370–4380. [CrossRef]
140. Tai, R.; Lin, L.; Zhu, Y.; Su, R. Privacy-preserving co-synthesis against sensor–actuator eavesdropping intruder. *Automatica* **2023**, *150*, 110860.
141. Kulkarni, Y.R.; Jagdale, B.; Sugave, S.R. Optimized key generation-based privacy preserving data mining model for secure data publishing. *Adv. Eng. Softw.* **2023**, *175*, 103332. [CrossRef]
142. Saleous, H.; Ismail, M.; AlDaajeh, S.H.; Madathil, N.; Alrabaee, S.; Choo, K.-K.R.; Al-Qirim, N. COVID-19 pandemic and the cyberthreat landscape: Research challenges and opportunities. *Digit. Commun. Netw.* **2023**, *9*, 211–222. [CrossRef]
143. Feng, Y.; Luo, Y.; Yang, J. Cross-platform privacy-preserving CT image COVID-19 diagnosis based on source-free domain adaptation. *Knowl.-Based Syst.* **2023**, *23*, 110324. [CrossRef] [PubMed]
144. Wang, Y.; Luo, Y.; Liu, L.; Fu, S. pCOVID: A Privacy-Preserving COVID-19 Inference Framework. In *Algorithms and Architectures for Parallel Processing: Proceedings of the 22nd International Conference, ICA3PP 2022, Copenhagen, Denmark, 10–12 October 2022*; Springer Nature: Cham, Switzerland, 2023; pp. 21–42. [CrossRef] [PubMed]
145. Dhasarathan, C.; Hasan, M.K.; Islam, S.; Abdullah, S.; Mokhtar, U.A.; Javed, A.R.; Goundar, S. COVID-19 health data analysis and personal data preserving: A homomorphic privacy enforcement approach. *Comput. Commun.* **2023**, *199*, 87–97.
146. Sei, Y.; Ohsuga, A. Private true data mining: Differential privacy featuring errors to manage Internet-of-Things data. *IEEE Access* **2022**, *10*, 8738–8757. [CrossRef]
147. Boubiche, D.E.; Imran, M.; Maqsood, A.; Shoaib, M. Mobile crowd sensing–taxonomy, applications, challenges, and solutions. *Comput. Hum. Behav.* **2019**, *101*, 352–370. [CrossRef]
148. Ota, F.K.; Meira, J.A.; Frank, R.; State, R. Towards Privacy Preserving Data Centric Super App. In Proceedings of the 2020 Mediterranean Communication and Computer Networking Conference (MedComNet), Arona, Italy, 17–19 June 2020; pp. 1–4. [CrossRef]

149. Wang, W.; Li, X.; Qiu, X.; Zhang, X.; Zhao, J.; Brusic, V. A privacy preserving framework for federated learning in smart healthcare systems. *Inf. Process. Manag.* **2023**, *60*, 103167.
150. Muthukrishnan, G.; Kalyani, S. Differential Privacy with Higher Utility through Non-identical Additive Noise. *arXiv* **2023**, arXiv:2302.03511. [CrossRef]
151. Majeed, A.; Hwang, S.O. Quantifying the Vulnerability of Attributes for Effective Privacy Preservation Using Machine Learning. *IEEE Access*. **2023**, *11*, 4400–4411.
152. Dina, A.S.; Siddique, A.B.; Manivannan, D. Effect of Balancing Data Using Synthetic Data on the Performance of Machine Learning Classifiers for Intrusion Detection in Computer Networks. *IEEE Access* **2022**, *10*, 96731–96747. [CrossRef]
153. Zhao, J.; Cheong, K.H. Obfuscating community structure in complex network with evolutionary divide-and-conquer strategy. *IEEE Trans. Evol. Comput.* **2023**. [CrossRef]
154. Pan, Y.L.; Chen, J.C.; Wu, J.L. Towards a Controllable and Reversible Privacy Protection System for Facial Images through Enhanced Multi-Factor Modifier Networks. *Entropy* **2023**, *25*, 272. [CrossRef]
155. Slavković, A.; Seeman, J. Statistical data privacy: A song of privacy and utility. *Annu. Rev. Stat. Its Appl.* **2023**, *10*, 189-218. [CrossRef]
156. Fu, N.; Ni, W.; Hu, H.; Zhang, S. Multidimensional grid-based clustering with local differential privacy. *Inf. Sci.* **2023**, *623*, 402–420. [CrossRef]
157. Chen, J.; Xue, J.; Wang, Y.; Huang, L.; Baker, T.; Zhou, Z. Privacy-Preserving and Traceable Federated Learning for data sharing in industrial IoT applications. *Expert Syst. Appl.* **2023**, *213*, 119036. [CrossRef]
158. Brunotte, W.; Specht, A.; Chazette, L.; Schneider, K. Privacy explanations—A means to end-user trust. *J. Syst. Softw.* **2023**, *195*, 111545. [CrossRef]
159. Stergiou, C.L.; Bompoli, E.; Psannis, K.E. Security and Privacy Issues in IoT-Based Big Data Cloud Systems in a Digital Twin Scenario. *Appl. Sci.* **2023**, *13*, 758. [CrossRef]
160. Dhirani, L.L.; Mukhtiar, N.; Chowdhry, B.S.; Newe, T. Ethical Dilemmas and Privacy Issues in Emerging Technologies: A Review. *Sensors* **2023**, *23*, 1151. [CrossRef]