IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 10 July 2023, accepted 18 July 2023, date of publication 21 July 2023, date of current version 27 July 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3297646

RESEARCH ARTICLE

When AI Meets Information Privacy: The Adversarial Role of AI in Data Sharing Scenario

ABDUL MAJEED AND SEONG OUN HWANG, (Senior Member, IEEE)
Department of Computer Engineering, Gachon University, Seongnam 13120, Republic of Korea

Corresponding authors: Seong Oun Hwang (sohwang@gachon.ac.kr) and Abdul Majeed (ab09@gachon.ac.kr)

This work was supported by the Institute of Information and Communications Technology Planning and Evaluation (IITP) Grant funded by the Korean Government [Ministry of Science and ICT (MSIT)] under Grant 2021-0-00540.

ABSTRACT Artificial intelligence (AI) is a transformative technology with a substantial number of practical applications in commercial sectors such as healthcare, finance, aviation, and smart cities. AI also has strong synergy with the information privacy (IP) domain from two distinct aspects: as a protection tool (i.e., safeguarding privacy), and as a threat tool (i.e., compromising privacy). In the former case, AI techniques are amalgamated with the traditional anonymization techniques to improve various key components of the anonymity process, and therefore, privacy is safeguarded effectively. In the latter case, some adversarial knowledge is aggregated with the help of AI techniques and subsequently used to compromise the privacy of individuals. To the best of our knowledge, threats posed by AI-generated knowledge such as synthetic data (SD) to information privacy are often underestimated, and most of the existing anonymization methods do not consider/model this SD-based knowledge that can be available to the adversary, leading to privacy breaches in some cases. In this paper, we highlight the role of AI as a threat tool (i.e., AI used to compromise an individual's privacy), with a special focus on SD that can serve as background knowledge leading to various kinds of privacy breaches. For instance, SD can encompass pertinent information (e.g., total # of attributes in data, distributions of sensitive information, category values of each attribute, minor and major values of some attributes, etc.) about real data that can offer a helpful hint to the adversary regarding the composition of anonymized data, that can subsequently lead to uncovering the identity or private information. We perform reasonable experiments on a real-life benchmark dataset to prove the pitfalls of AI in the data publishing scenario (when a database is either fully or partially released to public domains for conducting analytics).

INDEX TERMS AI-powered attacks, artificial intelligence, background knowledge, compromising privacy, data publishing, personal data, privacy, safeguarding privacy, synthetic data, utility.

I. INTRODUCTION

Data has replaced oil as the most economically desirable resource in the world, and therefore, most companies strive to acquire good data to accumulate profits. Data are no longer just raw materials but constitute a product with a wide range of financial opportunities. In the coming years, data can play a huge role in advancing science as well as influencing societies around the world. In some sectors, such as healthcare, a database is regarded as a living thing that offers multiple benefits to stakeholders (clinicians, patients, medical experts, manufacturers, etc.). Most companies invest in ways to draw value out of their data, especially by developing data-driven and data-centric tools. The ongoing pandemic was managed well in some countries that have access to good data and the best AI models [1]. In the AI-driven era, data can be used for multiple purposes, such as informed decision-making, reliable prediction, pandemic control, recommendations, and medical analytics.

Although data can contribute significantly in many domains and applications, there have been five major bottlenecks in data governance/use in both private and public sectors in recent times.

1) *Privacy preservation*: Safeguarding privacy while extracting enclosed knowledge.

The associate editor coordinating the review of this manuscript and approving it for publication was Derek Abbott.

2) *Quality*: Enhancing data quality without rigorous ETL processes and transformations.
3) *Responsible use*[1]: Deriving fair, transparent, impartial, equity-aware, and informed decisions without misusing the data.
4) *Ethical compliance*: Implementing fair information principles while ensuring privacy throughout the data life cycle.
5) *Black-box processing*: Empowering people by opening/explaining the data processing mechanism (or use), and communicating the risk of misuse.

Apart from other bottlenecks, privacy preservation has become a major obstacle amid the rapid rise in digital solutions (e.g., IoT) and learning paradigms (machine, deep, few-shot, split, etc.) [2], [3]. Amid the COVID-19 pandemic, privacy was regarded as one of the major bottlenecks when it came to the adoption/use of digital tools that acquire and use personal data[2] [4]. The practical approaches to preserving privacy are encryption, anonymization, pseudonymization, masking, and obfuscation. Anonymization has been widely used in commercial environments for safeguarding the privacy of personal data because of its low computing overhead, and it recently became law (e.g., mandatory) in some advanced countries. However, the invisible threats to anonymization mechanisms from AI tools remain unexplored in the literature.

A. MAJOR CONTRIBUTIONS

The major contributions of this paper to the information privacy field are demonstrated below.

- We discuss the synergy of AI with the information privacy domain and highlight the bright and dark sides of this synergy. Specifically, we present a brief overview of AI's role in the information domain from two aspects: preserving privacy and compromising privacy.
- We demonstrate the threats to individual privacy that have significantly evolved over time (even for anonymized data) from background knowledge (BK) and the surge in data on public repositories/social sites.
- We provide an extended taxonomy of BK including AI, specifically synthetic data (SD), which can be used to compromise privacy when carefully crafted (via generative models) whereas most of the existing methods do not take into account SD-based knowledge while anonymizing data. Recently, many generative models have been developed that can produce SD (tables, images, time series, etc.) of supreme quality which can constitute BK and can be used to re-identify people and to infer their associated sensitive information from the published data [5].
- We conducted experiments on real-world datasets to prove the feasibility of the proposed method in realistic scenarios. The experimental analysis confirms the findings of this paper and proves that good quality SD can lead to breaching an individual's privacy in some cases.
- To the best of our knowledge, this is the first work that regards SD as BK and highlights potential privacy threats stemming from it in data publishing scenarios. This extended knowledge can pave the way to rectifying privacy mechanisms, so they preserve privacy in an adequate way against present-day AI-powered attacks.

The remainder of this paper is organized as follows. Section II presents the preliminaries related to the subject matter discussed in this paper. Specifically, we present different types of attributes enclosed in personal data, privacy models, privacy-preserving data publishing (PPDP) processes, and state-of-the-art studies used for PPDP. Section III presents the synergies between AI and the privacy domain from two broader aspects. Section IV discusses BK as a major threat for compromising an individual's privacy. Section V describes the simulation results that were obtained from detailed experiments on the benchmark dataset to prove the significance of our proposed method in realistic scenarios. We conclude this paper in Section VI.

II. BACKGROUND AND LITERATURE SURVEY

In this section, we first present the background about the subject matter presented in this paper and then discuss the SOTA studies that have been proposed to safeguard as well as breach privacy in data publishing scenarios.

A. BACKGROUND

Privacy is all about hiding personal information from prying eyes (i.e., preventing public access). It is considered a fundamental human right and is imperative for individualism, autonomy, and self-respect. Privacy invasion can bring serious consequences to the victims, such as loss of respect, loss of a job, and social stigma. Therefore, most companies that utilize personal data pay ample attention to privacy protection. There are four main dimensions of privacy: information, communication, territory, and the body. This work fits into the first category that deals with personal data from multiple aspects (e.g., aggregate, store, process, anonymize, share, use, discard, etc.). Personal data can be contained in various formats (tables, time series, trace, matrix, etc.). For our purposes, this work assumes personal data are enclosed in tables.

Figure 1 presents a high-level overview of PPDP and the corresponding privacy threats that can still occur from the published data based on the BK possessed by an adversary. Specifically, an adversary uses information available from external sources to compromise privacy. This background knowledge has remained a serious threat to privacy and is increasing over time. To safeguard privacy, many models have been developed. Notable developments are: k-anonymity [6], ℓ-diversity [7], t-closeness [8], and differential privacy (DP) [9]. All these models have been rigorously upgraded and used in the PPDP process.

[1] https://redasci.org/
[2] https://cipesa.org/2021/08/data-protection-in-africa-in-the-age-of-covid-19/

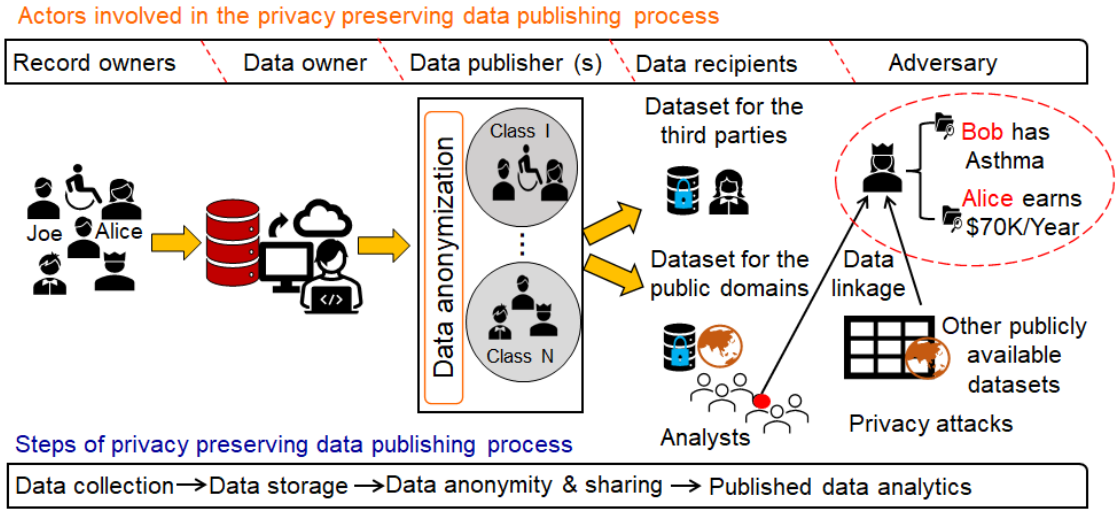


FIGURE 1. Overview of PPDP mechanisms and privacy attacks that can stem from published data by linking auxiliary (background) information.

B. LITERATURE SURVEY

Privacy preservation and breaching privacy are two important research topics that have been rigorously explored from multiple perspectives. Jia et al. [10] proposed a DP-based approach with a weaker definition of privacy to improve the utility of data analysis. The authors proposed a generic definition of privacy that can be widely used, and it has the potential to precisely estimate the privacy risks. Srijayanthi and Sethukarasi [11] developed a new clustering-based anonymization model to effectively preserve privacy and utility in data publishing scenarios. The proposed model performs dimensionality reduction and feature selection to lower the complexity of the anonymization process. Majeed and Hwang [12] developed a new anonymization method based on ML concepts. The authors improved various main steps of the anonymization process by using ML techniques. The ML-powered anonymization method significantly improved the privacy and utility results in data-sharing scenarios. Jha et al. [13] proposed a new anonymization scheme for stream data. The authors map the proposed z-anonymity to k-anonymity, and the proposed model can generically apply to any type of stream data. Chen et al. [14] combined the DP and k -median clustering to address the privacy issues in the data publishing without losing guarantees on anonymized data quality.

Recently, a clustering and t-closeness-based anonymization method has been developed to prevent similarity and skewness attacks in publishing multi-dimensional data [15]. The proposed method yielded better performance in terms of both privacy and utility. Hindistan and Yetkin [16] combined the DP and generative adversarial networks (GAN) model to protect sensitive data in industrial IoT. The proposed hybrid method can defend against contemporary privacy attacks with minimal loss of accuracy in data sharing. Tang et al. [17] proposed a DP-based mechanism for guaranteeing user privacy in the sequential publication of data. Similarly, a new model to measure the disclosure risks appropriately was recently developed [18]. The proposed measure can efficiently compute the disclosure risk in large datasets. Wang et al. [19] developed various privacy-preserving protocols for cloud-based inference systems for COVID-19 scenarios. The proposed protocols can provide higher security and privacy to outsourced data and inference results, respectively. Soliman et al. [20] proposed a framework for releasing time series data in raw form with the researchers while safeguarding the privacy of patients. Despite these approaches, it is still very challenging to eliminate all privacy risks amid the rapid rise in the data at auxiliary sources. In the next paragraph, we present SOTA approaches used to compromise privacy.

With the rapid proliferation of generative AI techniques, many de-anonymization (or privacy breaching) methods have also been developed that can compromise the individual's privacy. Sundaram et al. [21] devised a technique that has proved that attribute inference attacks can likely stem from randomly chosen records from SD. The authors have shown that a good quality SD can lead to privacy breaches in some cases. Hittmeir et al. [22] studied the similarities between SD and real data and proved that SD can lead to SA disclosure. However, only a few experiments were performed, and therefore, the conclusions are not general. Little et al. [23] compared the utility and privacy disclosures from the SD. This work also proves that SD can lead to privacy disclosure when some portions of the SD are very close to the real data. Ruiz et al. [24] developed a framework to empirically evaluate the privacy guarantees in SD, and suggested that disclosure risk in some parts of SD can be higher than in real data. Hittmeir et al. [25] developed a baseline to evaluate disclosure risk from SD. The authors assessed the likelihood of disclosure risk in Ensemble Vote and Radius Nearest Neighbor techniques. Yang et al. [26] also analyzed the disclosure

risk of the five most widely used anonymization methods. The authors concluded that anonymized data is subject to disclosure of various risks, and their findings can guide the data owners in choosing suitable anonymization methods and parameters to lower the re-identification risks. From the above analysis, it can be concluded that SD also entails various kinds of privacy risks. Hence, strong privacy-preserving methods are needed to secure personal data from malevolent adversaries in future endeavors.

III. WHEN AI MEETS INFORMATION PRIVACY

Just like any other field/discipline, AI has vastly influenced the information-privacy domain [27]. Specifically, AI can act as a defense tool–or an attack–in the privacy domain.

- AI as a defense/protection tool: AI can create synergy with the traditional anonymization mechanisms to effectively preserve privacy in PPDP. For example, Majeed and Hwang [28] used an AI-based anonymity approach to anonymize imbalanced data. Silva et al. [29] devised a privacy risk assessment framework to effectively preserve privacy by utilizing AI methods. Many AI-aided defense mechanisms have been developed thus far [27]. These developments highlight AI's role from a defense perspective in the privacy domain.
- AI as an attack tool: AI can be used to compromise privacy by creating synthetic data. Recently, AI has become a privacy-compromising tool in many respects, such as reconstructing parts of the data from anonymized data, in SA predictions, for membership inferences, and in QID estimation [30]. Ding et al. [31] proposed a method to re-identify people across social networks via AI. The authors re-identified up to 85% of the anonymized records. Since AI models can be fine-tuned with the help of hyperparameters, the chances of breaching privacy can therefore increase.

IV. BK: A MAJOR THREAT TO INFORMATION PRIVACY

In this section, we briefly introduce the methodology of this work, BK-based privacy breaches in data-sharing scenarios, and the proposed algorithm to address the potential threat posed by synthetic data.

A. METHODOLOGY

As pointed out by Yang et al. [26], inappropriate anonymization of personal data or flaws in the anonymization process often leads to privacy disclosures of various kinds including re-identification and private information disclosures. Most data owners gauge the privacy level in anonymized data before outsourcing the data to not lose the trust of the record owner. In some cases, adversaries can have very strong expertise in programming languages and can have access to datasets of various kinds, which serve as BK. By utilizing BK and programming expertise, adversaries can successfully compromise the privacy of some target individuals. With time, AI techniques are getting matured, which can also be leveraged to generate SD of good quality. Although all anonymization methods remove directly identifiable information as a recommended practice by laws and perturb the

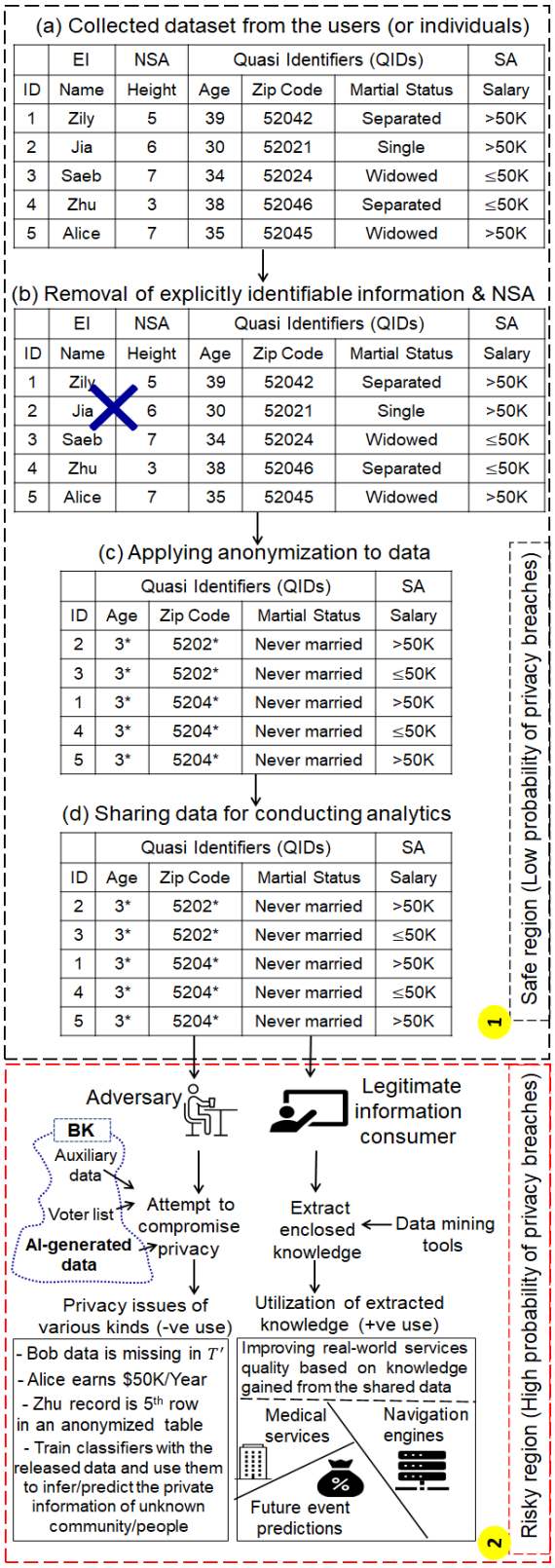


FIGURE 2. Methodology of the proposed work (data anonymization and privacy breaches leveraging BK including AI-generated data).

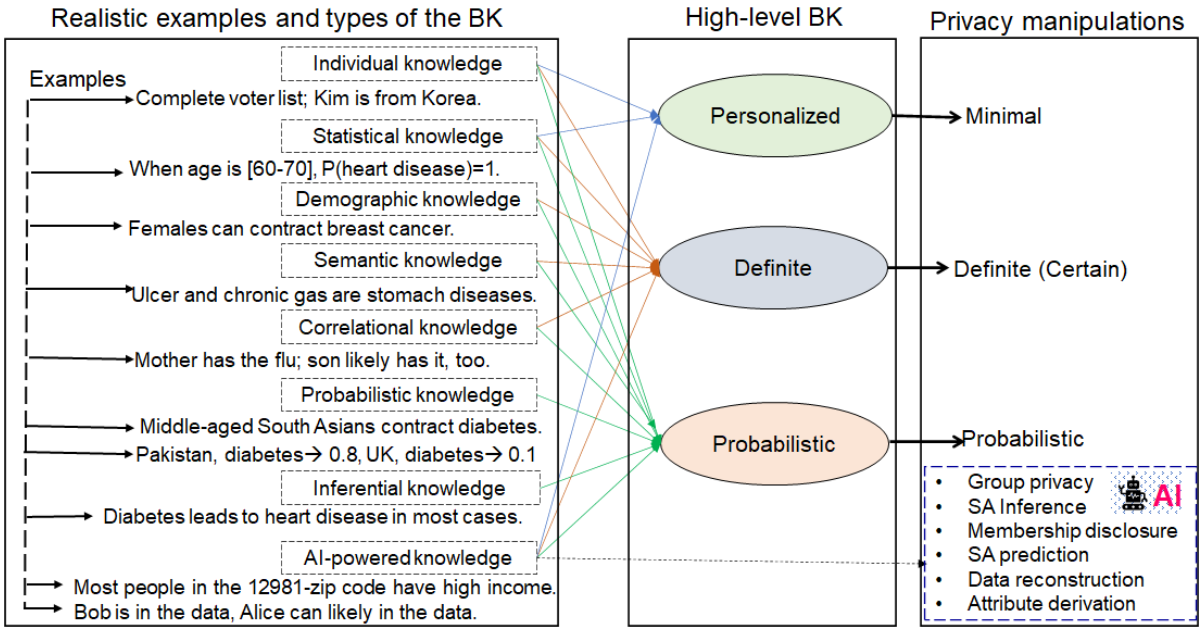


FIGURE 3. High-level taxonomy of BK (i.e., general and AI-based) and corresponding privacy manipulations.

structure of QIDs using generalization/suppression operations, data from some auxiliary sources can be acquired to perform linking, leading to privacy disclosures. Furthermore, the patterns extracted from the anonymized data can also lead to privacy disclosure in some cases.

Figure 2 demonstrates the methodology used in this work. As shown in Figure 2, personal data is anonymized before being shared with the data analyst. Specifically, the DIs and NSAs are removed from the data, and QIDs are generalized/suppressed. Subsequently, personal data in the anonymized form is outsourced for analytical or data mining purposes. The knowledge extracted from the anonymized data can be used in improving the quality of real-world services such as healthcare, navigation services, future event prediction, etc. Data sharing can foster the knowledge discovery process and can be very helpful in policymaking for the benefit of society. The values listed in Figure 2 (a) are exemplary values that are curated from real-world datasets to provide a comprehensive understanding of the anonymization mechanism and the possibility of privacy breaches subsequently. The type of attributes and corresponding values can vary from dataset to dataset. The values listed in Figure 2 (b) are the same as Figure 2 (a), but an initial step (e.g., removal of NSAs and EI) of the anonymity process is applied. The values listed in Figure 2 (c) are the anonymized version of the values given in the former tables. In Figure 2 (c), QIDs have been generalized using the generalization hierarchy of each QID. The values in Figure 2 (d) are the output of the anonymity process that can be readily shared with legitimate information consumers for conducting analytics. Specifically, Figure 2 is an extended version of Figure 1 along with actual values of the QIDs and SA. Unfortunately, a copy of the outsourced data can be delivered to adversaries, which can be linked with other sources of data (a.k.a auxiliary data) to infer the identities (or private information) of individuals. In some cases, adversaries can use this data in training classifiers, leading to the prediction of sensitive knowledge or private information of unknown communities/individuals. Soon, privacy attacks from AI techniques will likely become more evident, and therefore, more strong privacy methods are needed to guarantee privacy in data outsourcing scenarios.

In the past, adversaries usually relied on non-AI-based knowledge(e.g., voter lists, online repositories, data readily posted by an individual on social networks, public information of individuals on multiple social networks, factual information (ovarian/breast cancer can occur in females only), etc.). However, due to the rapid proliferation of AI techniques and tools, fine-grained data can be generated with AI tools which can constitute BK in some cases. In recent years, many AI-based techniques have been developed that can generate all three types of data such as structured, semi-structured, and unstructured data. Hence, the scale and nature of BK are drastically changing with time, and this AI-based knowledge can also be used along with other sources cited above as BK to breach an individual's privacy. The findings included in this article can provide the foundation for securing personal data from such emerging threats in the coming years.

B. PRIVACY BREACHES VIA BACKGROUND KNOWLEDGE

Background knowledge is certain (or generic) information about a person or group of persons that is owned by an adversary, gathered from multiple sources (the internet, social network profiles, research papers, online repositories, etc.), and that can eventually be used to compromise a person or a

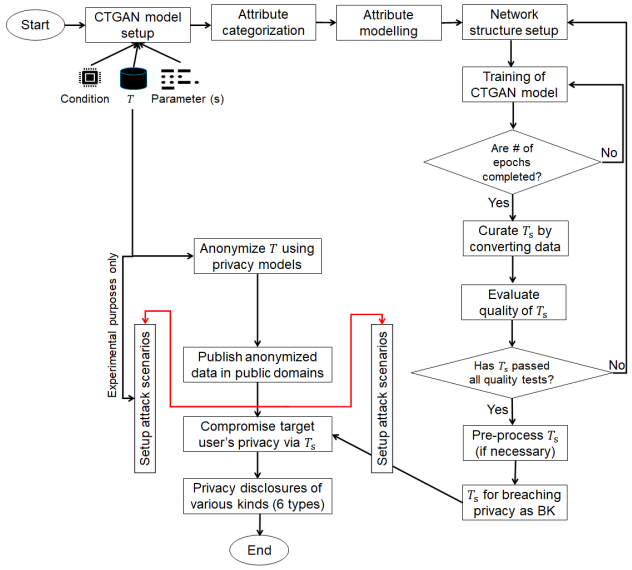


FIGURE 4. Flowchart of the proposed algorithm that was used to compromise individuals' privacy via T_S (e.g., AI-generated data).

group's privacy. Most privacy models take into consideration a certain form of BK while anonymizing data. BK can exist in multiple forms and can contribute to compromising privacy probabilistically, minimally, or definitely [32]. Figure 3 presents a high-level taxonomy of BK (extended from [32]) in the PPDP along with examples and corresponding privacy manipulations. To the best of our knowledge, none of the previous studies have highlighted AI as BK. As shown in Figure 3, AI can contribute to many categories of privacy breaches, requiring immediate attention from the research community to prevent misuse. The existing research has mainly focused on BK w.r.t. three types of privacy threats cited above, but AI has an advanced-threat landscape with far-reaching consequences for the general public. Therefore, futuristic privacy models must also consider AI-based BK to provide a strong defense against present-day AI-powered attacks.

The working procedure of the proposed algorithm that was used to compromise an individual's privacy by curating T_s of good quality via the CTGAN model is shown in Figure 4. There are two main components of the proposed algorithm: data generation (right side), and generated data utilization to compromise an individual's privacy (left side). The first component was implemented to curate T_s from real data, and the second component was implemented to leverage T_s as BK to compromise privacy from anonymized data. The details of curating tabular data by implementing the CTGAN model are given in the result section (see Fig. 7). To leverage T_s in breaching individuals' privacy, we devised six scenarios and performed statistical matching to determine the level of privacy breaches. The experimental details of scenarios and corresponding privacy disclosure results are reported in the relevant subsections (see subsections V-A-V-F). In our algorithm, we used some information from real data as well to justify our findings. For example, we devised sensitive rules from real data and then used them in quantifying the privacy breach level by using anonymized data and T_s, respectively.

C. PROPOSED ALGORITHM TO ADDRESS THE POTENTIAL THREAT POSED BY SYNTHETIC DATA

In this subsection, we propose an algorithm to address the potential threat posed by synthetic data in compromising an individual's privacy. The working procedure of the proposed algorithm used to address the potential threat posed by synthetic data is given in Figure 5. There are three main modules of the proposed algorithm: synthetic data generation using generative AI models, AI and non-AI BK-aware anonymization of real data, and anonymization data sharing for analytics while preserving the privacy of individuals. The concise details of each building block are given below.

1) SYNTHETIC DATA GENERATION USING GENERATIVE AI MODELS

In this module, SD is generated by mimicking the properties of real data. In recent years, plenty of generative AI methods have been proposed to either curate data of good quality or to optimize the synthetic data generation process. In the proposed method, we generate SD using the conditional tabular generative adversarial network (CTGAN) to use it as BK to quantify the strengths of the anonymization algorithm [33]. The CTGAN model has a condition that enables accurate modeling of all values as well as their distributions. It has been widely used to generate data of good quality for downstream tasks (e.g., training data development, data augmentation, privacy protection, etc.) It is important to note that some data owners release SD as is with the information consumers under the assumption that SD is an impure form of real data. This work tends to highlight that SD can pose privacy risk if it is very close to real data or combined with external sources (a.k.a auxiliary data). Due to the rapid rise in generative AI tools and data availability, the notion of privacy attacks has changed manifolds. Hence, a strong defense against BK that was not very common in the recent past is imperative.

2) AI AND NON-AI BK-AWARE ANONYMIZATION OF REAL DATA

In this module, the anonymization of data is performed by rigorously considering both AI and non-AI BK to provide a strong defense against breaching an individual's privacy. Seven key steps in the proposed algorithm are sequentially applied to generate anonymized data from T. The objectives and methods employed in each step are given below.

In the data cleaning step, data is cleaned with the help of sophisticated pre-processing techniques. The main objectives of this step are to make the data interpretation easier and to prevent wrong conclusions from the anonymized

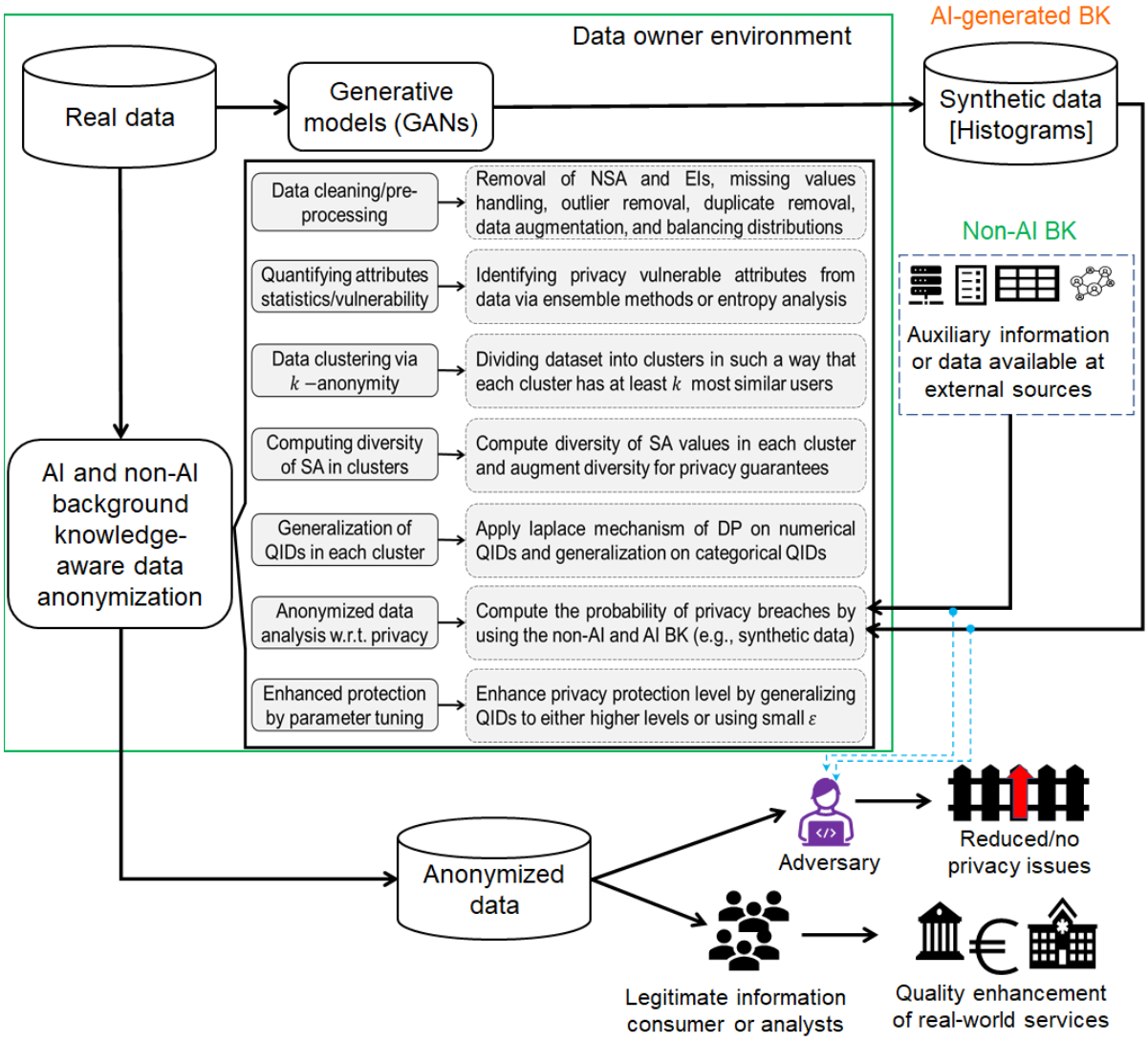


FIGURE 5. Conceptual overview of the proposed algorithm to address the potential threat posed by synthetic data (e.g., AI-generated knowledge).

data. In addition, the data is cleaned to prevent difficulty in anonymizing it. For example, if there are outliers or missing values in the data, they can lead to either wide intervals or improper generalization. In some cases, the deletion of records encompassing missing values can lead to a reduction in data size, and therefore, it is better to impute them with the mean of the entire column. In any real-world T, there are four types of attributes: sensitive attributes (SA), quasi-identifiers (QIDs), non-sensitive attributes (NSAs), and explicit identifiers (EIs). The details of all four attribute types along with the examples are given in our recent work [34]. In the beginning, we remove two types of attributes from the data to prevent identity disclosure and to lessen the computing complexity as standard practice in PPDP [35]. The remaining two attributes are QIDs and SA, respectively. We denote QIDs with set Q, where $Q = \{q_1, q_2, \ldots, q_p\}$, and SA with Y. The overall structure of T is given in Eq. 1.

$$T = \begin{pmatrix} x_i & q_1 & q_2 \cdots & x_p & Y \\ x_1 & v_{q_1}^{x_1} & v_{q_2}^{x_1} \cdots & v_{q_p}^{x_1} & y_1 \\ x_2 & v_{q_1}^{x_2} & v_{q_2}^{x_2} \cdots & v_{q_p}^{x_2} & y_2 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ x_N & v_{q_1}^{x_N} & v_{q_2}^{x_N} \cdots & v_{q_p}^{x_N} & y_{p'} \end{pmatrix} \quad (1)$$

In Eq. 1, q_1, q_2, q_p can be used to denote age, gender, and race, respectively. Y can be used to denote income/disease. In real-life T, each QID has a different cardinality depending upon its unique values. After giving an understandable structure to T, we apply pre-processing techniques to T. Specifically, our algorithm imputes missing values rather than deletion to maintain the data size. The missing values in the numerical columns are imputed with the mean of the respective column. In contrast, the missing values in the categorical column are

substituted with under-representative values in the respective columns. The proposed algorithm removes outliers using the $min - max$ approach and domain analysis. The duplicate records are removed using the similarity information between records. We assume that records that are located next to each other, and all values are identical are duplicate records. Our algorithm removes one tuple/record from the duplicate records. If the quality of T is poor, augmentation and distribution analysis can be applied to improve data quality before anonymization. In the second step, statistics regarding the vulnerability, utility, information gain, etc. of QIDs are computed. The main purpose of computing these statistics is to prevent privacy disclosures or utility issues in published data analytics [36], [37]. To compute vulnerability, we employed an ensemble method named, random forest [38]. Random forest (RF) is one of the reliable machine learning methods with a substantial number of applications. We employed the RF method to rank the QIDs from the perspective of vulnerability. We build the model with original and shuffled data (e.g., the value of one QID is column-wise permuted). The accuracy was analyzed before and after data shuffling and the vulnerability of the QIDs were computed. With the help of the RF-based method, QIDs that are riskier in terms of privacy were identified. By determining the vulnerability information, ample attention can be paid to vulnerable QIDs during their generalization.

In the third step, T is divided into non-overlapping clusters of size k (i.e., the minimum size of each cluster is k). The total # of clusters given the size of the T can be computed using Eq. 2.

$$C_n = \frac{N}{k} \tag{2}$$

where $N = |T|$ denotes size of the T, and k is a privacy parameter.

While assigning records to clusters, we compute similarity $\mathcal{S}$ between records using Eq. 3. The $\mathcal{S}$ value between two records, x_a and x_b, can be computed using Eq. 3.

$$\mathcal{S}(x_a, x_b) = \frac{\sum_{i=1}^{p} x_{a_i} \times x_{b_i}}{\sqrt{\sum_{i=1}^{p} x_{a_i}^2} \times \sqrt{\sum_{i=1}^{p} x_{b_i}^2}} \tag{3}$$

where i represent the QID present in data, and p denotes the total # of QIDs.

In the clustering process, records are mapped into clusters based on the $\mathcal{S}$ value as well as the k-anonymity criteria, meaning each cluster has at least k records. Our algorithm generates compact clusters, leading to lower generalization intervals during anonymization. By having identical records in each cluster, the information loss is restrained. After creating clusters, uncertainty $\mathcal{U}$ is computed from the SA column. The $\mathcal{U}$ is computed with the help of the Shanon entropy concept using Eq. 4.

$$\mathcal{U}(C_i) = -\sum_{i=1}^{|Y|}[(p_i) \times ln(p_i)] \tag{4}$$

where p_i represents the proportion of each SA value in a class C_i. The $\mathcal{U}$ value ranges between 0 and 1 (i.e., $\mathcal{U} \in [0, 1]$).

For $k = 3$, if all users in a class share the same SA value (e.g., ≤ 50 K), then $\mathcal{U}(C_i)$ will be zero. Similarly, For $k = 3$, if two users have income higher than 50 K, and one user has less than 50 K, then $\mathcal{U}(C_i)$ will be 0.91. With the help of the $\mathcal{U}$ concept, ample attention can be paid to classes having low $\mathcal{U}$ to control privacy breaches.

In the next step, the original values of the QIDs are replaced with generalized values. To convert QID values, we applied two distinct mechanisms: Generalization hierarchy and laplace noise addition. The laplace mechanism is highly suitable for anonymizing numerical data [14]. The anonymized numerical QID (q') can be obtained using laplace mechanisms as expressed in Eq. 5.

$$q' = q + n \tag{5}$$

where $n \sim \text{lap}(\frac{\Delta F}{\epsilon})$ is a randomly generated noise and it can be drawn from laplace distribution with scale factor $\frac{\Delta F}{\epsilon}$. Figure 6 demonstrates the overview of both mechanisms that were used to anonymize data. By applying both these mechanisms at the same time, privacy-preserved anonymized data is curated for further analysis.

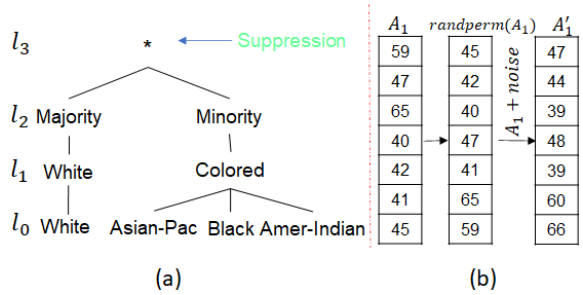


FIGURE 6. Overview of the generalization hierarchy and laplace noise addition method used in converting real values of QIDs into anonymized form.

In the next step, anonymized data is evaluated w.r.t privacy level before releasing it to researchers/data miners. In the proposed algorithm, both non-AI and AI-based BK are considered to quantify the level of privacy in anonymized datasets whereas existing methods only take into account the non-AI BK. By considering only non-AI knowledge, strict privacy guarantees cannot be achieved in data publishing scenarios, and the probability of explicit privacy breaches is high when the adversary has access to some high-quality SD. During the evaluation of privacy strengths, multiple records as a BK are extracted from auxiliary sources and SD, and their privacy disclosure was performed from anonymized data. Furthermore, some other useful knowledge (e.g., minority values, common patterns, frequency of values, etc.) derived from the SD was also leveraged to determine the privacy status of individuals in anonymized data. If the privacy disclosures are high in anonymized data, then more strict values for the privacy parameter (e.g., k, ϵ, etc.) were applied to increase the defense level. Once all privacy tests are passed, then data

can be outsourced for knowledge discovery purposes. In some cases, SD can also be amalgamated with the real data to improve distribution skewness or less # of records problems, leading to better anonymization of data.

Effectiveness of proposed algorithm in terms of privacy and utility: The proposed algorithm can effectively maintain the balance between utility and privacy because the users are grouped in the classes based on similarity, and diversity in the SA column is also considered while anonymizing data. The first method ensures that generalization is performed to lower levels of the hierarchy in most cases to effectively preserve the semantics and truthfulness of real data in anonymized data, leading to the better utility of data. Privacy is ensured by applying laplace noise addition to numerical QIDs and hierarchy-based generalization of the categorical QIDs.

Comparisons of the proposed algorithm with the existing methods: The proposed algorithm is an extension of our previous algorithms [12], [28] with minor modifications, and therefore, it can better safeguard privacy without losing guarantees on data utility. The proposed algorithm can be compared in terms of effectiveness and efficiency with the existing SOTA methods by varying k and assuming % of data exposed to the public domains. The proposed method can yield better results because useful knowledge concerning the data composition is extracted with the help of the ML technique, and considered at the time of anonymizing data. By combining traditional and multidisciplinary approaches, the proposed algorithm can outperform the existing methods from the perspective of effectiveness and efficiency.

Application of the proposed algorithm to large datasets: The proposed algorithm has been applied to the adult dataset which is relatively large and encompasses attributes of mixed type (i.e., categorical and numerical). The proposed algorithm can also be applied to large datasets encompassing many attributes and records by choosing appropriate parameters for RF, k value, diversity criteria, choosing the optimal value of ϵ, and constructing generalization hierarchies of new QIDs. In the anonymization domain, some attributes (e.g., NSA and EIs) are deleted from the data at the start, and only a small subset of attributes is retained for further processing. Therefore, the computing complexity of the proposed algorithm does not rise drastically. In addition, some parts of the data need either low or no anonymization, owing to general patterns in them [39]. Therefore, the computing complexity is not very high in most cases. However, the computing complexity of the proposed algorithm can still rise with the vertical and horizontal expansions in the data in real-world cases. The parallel implementation and low-cost operations can be integrated with the proposed algorithm to reduce computational complexity while anonymizing large datasets.

Limitations of the proposed algorithm: The current implementation of the proposed algorithm can work well in the single SA (e.g., disease/income) scenario only. Hence, it cannot be directly applied to scenarios where data encompasses multiple SAs (e.g., disease and income, income and political views, etc.). When the underlying data to be anonymized is poisoned, our algorithm may not yield reliable results from the perspective of data mining and knowledge discovery. The utility and privacy results can be low when real data is skewed, noisy, and incomplete. The computing complexity can increase with changes in data size (e.g., row-wise and column-wise). It can only be applied to the datasets that have already been curated from the relevant users/individuals. Further, it may offer less protection w.r.t. group privacy. In real-time processing paradigms (e.g., cloud environments) our method may face some challenges such as the highly customized implementation of libraries needed for computing statistics about T, interactive interface development for query acquisition from analysts, and visualizations of query answers for robust analysis without jeopardizing user's privacy. However, most of the above-cited issues can be resolved by creating synergy between our algorithm and the DP models in future research. Lastly, some more AI methods (e.g., gradient boosting machines) can be integrated with our algorithm to identify the intrinsic characteristics of data, leading to effective resolution of privacy and utility in challenging scenarios. In addition, we intend to explore data balancing and noise removal methods to prepare sound data before applying anonymization to address these limitations in future research.

Applicability of the proposed algorithm to different data modalities: The proposed algorithm was primarily designed for tabular data environments, where multiple attributes act as QIDs, and one as SA. The proposed algorithm can be applied to other data modalities such as text and images after modifications. For example, it can be applied to image data by concealing the parts that likely reveal someone's identity or adding more noise in less class-independent features [40], [41]. Our algorithm can also be applied to text data by introducing a modification in the generalization part (e.g., synonyms-based generalization), altering the structure of data, and making it identical to tabular data. Also, some modifications in the pre-processing part are needed depending on the data modality. Some of the DP-based methods have already been applied to text data [42], and therefore, we expect our algorithm can also be applied to both data modalities by making some changes in the relevant parts.

Potential real-world applications of the proposed algorithm: The potential real-world applications of the proposed algorithm are discussed below.

- It can contribute to making data available at a large scale, which, in return, can contribute to conducting innovative research and validating various research hypotheses. In most real-world applications, good data is imperative in extracting knowledge and improving the quality of existing services such as medical applications, event prediction, discovering treatment methods, etc. The existing anonymization methods distort the data too much, which

can impact the quality of decisions. In contrast, the proposed algorithm performs only minimal generalization and effectively solves the problem of data availability and quality, leading to better analytics and data mining.

- It can be used as a pipeline with the ML classifiers to solve the privacy and utility issues in the training process [43]. For example, most ML models memorize the data during training which can lead to privacy disclosure at inference time. Also, some anonymization models destroy the quality of data which can yield deficient accuracy in ML models. To this end, our algorithm can be used to reduce privacy risks and utility issues in ML.
- It can be highly useful in anonymizing medical records and bank data which are vital for researchers and analysts. In the recent pandemic, the need for data sharing in the early phases of the pandemic was crucial to understanding the dynamics of this disease. Similarly, the anonymization and sharing of rare/dangerous disease data are vital to foster treatment methods. To this end, our algorithm is a better candidate than the former methods due to its ability to strike the balance well between two competing goals (e.g., privacy and utility).
- It can provide sufficient resilience against generative AI-based attacks on anonymized data. As a result, both data owners and providers can be less worried about privacy disclosures. To the best of the authors' knowledge, none of the previous studies considered the SD-based BK, and therefore, they are vulnerable to privacy disclosures in the presence of AI-based BK. In contrast, the proposed algorithm can better safeguard users' privacy and can contribute to lowering people's hesitations and concerns related to their personal data privacy.
- The proposed algorithm can contribute to drawing fair and undisputable decisions from the data by preserving the semantics of real data to the extent possible. For example, it abstracts some general patterns and does not apply anonymity to them, and DP with proper privacy budgets is utilized to prevent the distortion in data.
- It can be applied to different scenarios involving similar structures of the data. For example, our work can be applied to set-valued data encompassing transaction data with minor modifications. Hence, it can contribute to marketing, customer segmentation, and recommendation systems.
- It can be applied to privacy preservation in web data analytics. The web data encompasses a rich source of information for marketing purposes, and user behaviors analysis with strict privacy guarantees is paramount [44]. To this end, our algorithm can be applied to hide explicit identity information while permitting informative analysis of web data.
- Recently, bias mitigation in AI applications has become a very hot issue, and anonymization is one of the pertinent solutions [45]. To this end, our algorithm can produce good-quality data that can lower the possibility of bias in AI applications.

The evaluation of the proposed algorithm in the above-cited applications can be performed with the help of established metrics. For example, the evaluation of utility can be performed using special purpose (e.g., accuracy, precision, recall, F_1 score, etc.) and general purpose (e.g., information loss, distortion, semantic loss, etc.) metrics. Privacy can be evaluated with the help of multiple metrics such as disclosure risk, probabilistic disclosure, re-identification probability, Sa disclosure, indistinguishability, etc. The quality of the decision drawn from anonymized data can be evaluated with the help of accuracy, true positive rate, etc. Similarly, the performance in other applications can also be evaluated using relevant metrics. Recently, some open-source tools have also been proposed to evaluate the privacy, utility, and other objectives of the anonymity solutions [46]. In some cases, a single evaluation metric was used to measure the performance in more than one application at the same time [47].

Adaption of the proposed algorithm to handle different types of privacy threats: The proposed algorithm can effectively solve the identity and SA disclosures in the relational data. The protection against these threats is ensured by assembling k-records with identical QID values, and higher diversity in the SA column. However, its amalgamation with the DP can contribute to handling different types of privacy threats, such as inference attacks or membership attacks in AI environments [48], [49]. Furthermore, our algorithm can be adopted to combat different types of advanced privacy threats by curating fused data (e.g., real data + synthetic data) using generative AI models [50], [51]. The joint use of data balancing and the DP model with the proposed algorithm is expected to offer a solid defense against different types of privacy threats, such as inference and membership attacks.

3) ANONYMIZED DATA SHARING FOR ANALYTICS WHILE PRESERVING THE PRIVACY OF INDIVIDUALS

In this module, anonymized data is shared with the analysts, researchers, and/or data miners for extracting useful knowledge from data without compromising individual privacy. In the proposed algorithm, rigorous privacy checks are performed on the anonymized data, and therefore, the probability of privacy breaches is low. The proposed algorithm considers SD as BK during data anonymization, and therefore, the probability of privacy risk can be restrained in anonymized data. The proposed algorithm can provide a better safeguard in the presence of AI and non-AI BK whereas the existing method only considers non-AI BK, leading to higher privacy breaches. Lastly, our algorithm exploits the benefits of differential privacy and generalization hierarchy to retain higher knowledge for information consumers. By exploiting intrinsic characteristics of attributes in T, our algorithm retains better semantics of T in T', leading to higher utility and privacy.

It is worth noting that any anonymization method with one additional step (e.g., consideration of AI-based BK which might be available to the adversary) can be applied to provide

better defense against AI-powered attacks. Our algorithm is generic and can be applied to provide a solid defense against both AI and non-AI-based BK in data-sharing scenarios.

In the proposed algorithm, different new and established algorithms/criteria have been implemented to accomplish the task of privacy preservation (against both BK and non-AI BK attacks) and utility enhancement in anonymized data. For example, two types (e.g., EIs and NSAs) of attributes were removed from the data based on the well-established criteria for PPDP. Similarly, missing data were handled with a newly developed data engineering pipeline to not lose data size as well as truthfulness. The vulnerable attributes were identified using the random forest method along with the data shuffling strategy [28]. Specifically, we build an RF model with unaltered and altered data to choose the QIDs that are more prone to identity disclosure from anonymized data. Later, we applied the k-anonymity model to create classes from the data in such a way that each class encompasses at least k users. By applying the k-anonymity model, the possibility of identity disclosure is restrained, and the probability of identifying any user from the data becomes $1/k$. Later, we applied the Shanon index method to compute and analyze the uncertainty regarding the SA value in each class. This index is reliable and efficient and widely used in the forest for quantifying the diversity of species. We applied this method to analyze the diversity of SA values in each class. If the value from this index is zero, then SA disclosure is certain (e.g., 100 %). The higher value ($\sim$1) is desirable to safeguard the SA disclosure in data-sharing scenarios.

The data generalization was performed with the help of two algorithms. The generalization of categorical QIDs was performed using generalization hierarchies. The anonymization of numerical QIDs was performed using the laplace method of the DP. Later, We analyze the anonymized data w.r.t privacy by using synthetic data created with the CTGAN model and some non-AI knowledge. Specifically, we performed statistical matching under different conditions and analyzed the corresponding privacy leakage. If privacy is well safeguarded (e.g., privacy disclosure is in the acceptable range), then anonymized data is considered final and can be outsourced. If the privacy disclosure rate is higher, we apply the stricter values of privacy parameters to enhance the protection level of privacy against contemporary privacy threats. In the end, privacy-preserved data is curated and outsourced for conducting analysis or data mining purposes. The implementation of the algorithms has been performed in Python (SD generation), *R* programming (QID's vulnerability computing), and Matlab (the rest of the steps of the proposed algorithm).

V. RESULTS AND DISCUSSION

To prove the adversarial role of AI in data publishing, we conducted exhaustive experiments on a real-life dataset named Adults.[3] The adult is a benchmark dataset that has been widely used to prove the feasibility of anonymity methods. It contains 32,561 records, mixed attribute types (e.g., numerical and categorical), and diverse values for SAs. We present the concise details (e.g., category of the attributes, attributes labels, type, and cardinality) of this benchmark datasets used in experimental evaluation in Table 1. In Table 1, C and N refer to categorical and numerical, respectively.

We pre-processed (e.g., we removed the incomplete tuples from the datasets, eliminated the outliers using min-max analysis and exploratory data analysis, and made the values consistent in each column) this dataset before utilization in experiments. Furthermore, we eliminated the redundant records that are adjacent to each other and have identical values row-wise. The experimentation with an error-free dataset allowed us to precisely quantify the number of privacy disclosures.

TABLE 1. Details of the real-life dataset used in the experiments.

Attribute category	Details of attributes		
QIDs/ SA	Attribute label	Cardinality	Type
QIDs	Gender	2	C
	Age	74	N
	Country	41	C
	Race	5	C
	Occupation	14	C
	Work class	8	C
	Education	16	C
	Education #	16	N
	Relationship	7	C
	Marital status	6	C
	Capital loss	118	N
	Capital gain	91	N
	Hours-per-week	93	N
	fnlwgt	21,648	N
SA	Income	2	C

Table 2 presents the information on SA encompassed in the adult dataset. The distribution of SA values is imbalanced, and one value occurs with a higher frequency than the other. The imbalanced distribution in SA poses serious threats to the applicability of some anonymization models (i.e., ℓ-diversity), and many records can be exposed to adversaries owing to no diversity in the SA column. Therefore, it is paramount to assess the quality of data before its anonymization to lower the probability of privacy breaches.

TABLE 2. Statistics of the SA's values present in the adult dataset.

Total records	Unique value of SA	Frequency in T
32,561	<=50 K	24, 720
	>50 K	7, 841

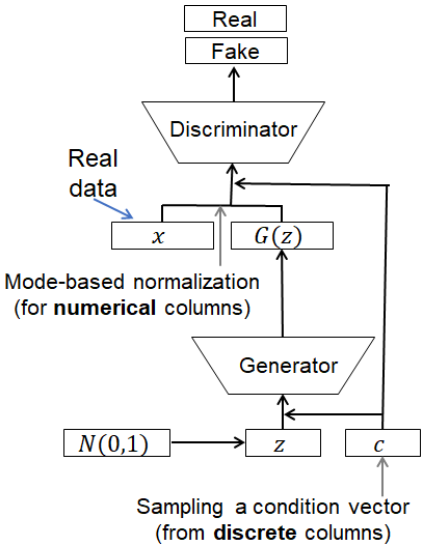
We implemented a CTGAN model to get a copy of SD for our BK-related analysis. The practical architecture and interface of the CTGAN model are given in Figure 7. Figure 7 (a) is about the working procedure of the CTGAN model that is used to generate SD of good quality, and Figure 7(b) shows the corresponding implementation in Python language. For example, sampling a conditional vector is one of the modules in the CTGAN model to ensure diversity in the generated

[3]http://ctgan-data.s3.amazonaws.com/census.csv.gz

data (7(a)), and Figure 7(b) shows the name of categorical columns that were used as a conditional vector in actual implementation to ensure diversity in the generated data. Similarly, the acquisition of real data is consistent in both figures. It is worth noting that Figure 7(b) demonstrates only the initial interface of the CTGAN model and other modules are implemented in separate files. As shown in Figure 7 (b), the input to the CTGAN model from the main interface is the URL of the real data, condition vector, epoch's count, and # of synthetic records to be generated. However, some parameter values were given as input through supportive files encompassed in the same folder. The output of the CTGAN model is synthetic data of the best quality. The CTGAN is a SOTA approach to SD generation because it imposes conditions on discrete columns, and mode normalization of numerical data, to keep functional relationships similar between synthetic data (T_s) and real data (T). For the experiments, we obtained a T_s of the best quality via the CTGAN.

Generating high-quality T_s is very challenging particularly when T is of low quality (i.e., skewed distributions, missing values, outliers, fewer records, etc.). Some GAN-based methods show poor performance in generating numerical columns, and they only generate categorical columns [52]. Some models yield poor performance when the size of real data is small or there exist multiple classes rather than binary class [53]. In this work, we used the CTGAN model with modifications to curate T_s of better quality. Furthermore, we pre-processed the T before generating the T_s to lower computing overheads and unnecessary operations. We used two parameters to curate T_s of good quality. The conditional vector was applied to categorical columns to correctly replicate all values in T_s from T. Without a conditional vector, the CTGAN model only learns values of high frequency, and less frequent values remain unlearned. Secondly, we modeled the distributions of the numerical columns using the VGM model because the values of numerical columns cannot be bounded in [-1, 1] form [54]. Furthermore, the training process was made more stable using the optimal loss function and PacGAN strategy to achieve T_s of good quality [55]. It is important to note that it will be hard to achieve the high quality T_s for a different dataset. For example, if all features in a real dataset are either categorical or numerical, then curating T_s of high quality can be hard. Similarly, if the size of T is very small, then curating T_s with higher diversity can be challenging, leading to poor quality T_s. Furthermore, if T is unlabelled, then curating T_s with sufficient accuracy is hard [56]. Some generative models convert real data into another form while generating T_s, which can be time-consuming and hard when the data size is large [57]. Lastly, generating T_s of high quality when T is encompassed in different modalities (i.e., tables, sensor readings, time series, etc.) can be hard owing to different formats and compositions of data.

Next, we highlight privacy risks stemming from SD when it is used as BK. The feasibility of the proposed idea was experimentally verified through six different aspects. (i) disclosure of SA distributions, (ii) group privacy disclosure, (iii) SA inference/disclosure, (iv) Data reconstruction attack, (v) SA prediction attack, and (vi) re-identification attack. To the best of the authors' knowledge, none of the previous studies have evaluated the privacy risks of AI-generated data from a broader perspective. The evaluation of SD from a broader perspective highlights the pitfalls of SD in the modern AI-driven era. While conducting experiments regarding privacy breaches, we used T_s, T, T', and non-AI BK as an input. In some cases, we generated sensitive rules, target values in which the attacker might be interested, and auxiliary data to quantify the amount of private information leakage. The output was the disclosure risk, classes at risk, disclosure of private information, success in data reconstruction, etc.



(a) Implementation architecture of the CTGAN.

```
from ctgan import CTGANSynthesizer
from ctgan import load_demo
# Acquisition of the real data.
data = load_demo()
# Specifying the columns that are categorical.
discrete_columns = [
    'workclass',
    'education',
    'marital-status',
    'occupation',
    'relationship',
    'race',
    'sex',
    'native-country',
    'income']
ctgan = CTGANSynthesizer(epochs=100)
ctgan.fit(data, discrete_columns)
print(data)
# No. of records to be included in the synthetic data
samples = ctgan.sample(32561)
# Writing data on file for downstream tasks.
with open ('synthetic_data.xxx','w') as f:
 f.write(samples.to_string(header=False, index=False))
print (samples)
```

(b) The CTGAN interface: T_s creation from T.

FIGURE 7. Implementation of the CTGAN model for T_s creation by correctly mirroring the properties of the T.

A. DISCLOSURE OF SA DISTRIBUTIONS

In the first analysis, we compared distributions of SA values in the T as well as T_s. We considered Profession and Income as SAs for BK-related analysis. The objective of analyzing the distributions was to verify the reflection of all values, as well as their distributions, from T to T_s. A comparative

analysis of value distributions for Profession (an SA) is shown in Figure 8. Referring to Figure 8, the black line shows the distribution of SA values in T, and the blue line shows the distribution of SA values in T_s. From Figure 8, we can see that the distributions of values are very close in T_s and T. These results and analysis indicate that good quality T_s offers valuable hints about the distribution and trends of values that can be exploited to compromise an individual's privacy. The T_s can allow adversaries to observe commonalities and differences among SA values, that, in return, can be used to infer the identity or SA information of some target individuals.

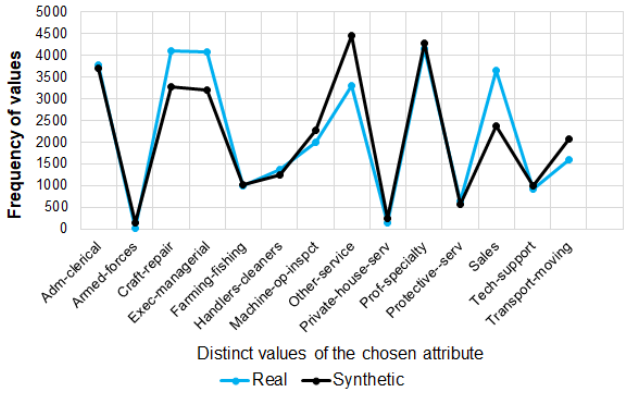


FIGURE 8. SA value distributions in T and T_s(the narrow gap indicates better quality of T_s).

With the help of Figure 8, we demonstrate the closeness between T and T_s that can lead to privacy breaches in some cases when combined with another type of BK. Similarly, for the other SA (Income), the difference between distributions was not sufficiently large. There were two distinct values for Income: $>$ 50 K and $\leq$ 50 K. The former had the lower frequency in both T and T_s, and the latter had the higher frequency in both datasets. However, the differences in frequency were not more than 10%. By knowing an SA label or QID information, the privacy of an individual can be compromised. These results confirm that SD can offer valuable information to adversaries regarding the distribution of SAs, which can lead to privacy disclosure through matching (or prediction).

B. GROUP PRIVACY DISCLOSURE

We observed that major categories in QID values were retained, as is, in the T_s produced by the CTGAN, and therefore, the chance to violate group privacy (GP) can occur based on those values. For example, race value white occurred 27,816 times and 24,139 times in T and T_s, respectively. We classified dominant QID values into two categories (super-major and major) and, in Table 3, report group privacy disclosures based on dominant QID values.

The values listed in Table 3 were obtained through co-relation between T and T_s. Through value-driven analysis of categorical QIDs present in T, we identified unique values and computed their frequencies. Subsequently, we classified them into major and super-major categories based on their frequencies. Later, we applied the same procedure to T_s and identified the major and super-major values and their frequencies. Lastly, we performed the analysis and computed the group privacy disclosures that can likely emanate from the released data. Through this analysis, we intend to highlight that if the generative model is good, the functional relationships between T and T_s can be higher, leading to group privacy breaches in some cases via the dominant QID values.

TABLE 3. Group privacy disclosure using SD as BK.

Category status	Dominant QID values	Group disclosure (%)
Super major	White , US	85.42%
Major	Black, M, P, G, C, I	9.59%

Abbreviations: US(United States), M(Mexico) P(Philippines), G(Germany), C(Canada), I(India).

C. SA INFERENCE/DISCLOSURE

We employed a disclosure-risk metric (a.k.a. re-identification probability) in two attack scenarios (BK and linking) to assess the performance w.r.t SA inference. In the BK attack, the adversary already has access to a real table and some facts about target users. In a linking attack, an adversary might try to link published and auxiliary data to match and infer the SA of individuals. The disclosure risk, D_{risk}, can be computed using Eq. 6:

$$D_{risk}(u_i, v_i) = \sum_{i=1}^{|b|} v_i / \sum_{m=1}^{k} v_m \tag{6}$$

where v_i is the SA value an adversary wants to infer for a target user, u_i, and the denominator shows the sum of the frequency of distinct SA values in a group. $D_{risk} = 1$ only when all users in a group share the same SA. We considered worst-case scenarios while measuring and comparing D_{risk} values from T and T_s.

We found correct matches based on QID values in different classes from T and T_s and computed the probability of income being either $>$50 K or $\leq$50 K by using multiple forms of the BK attack. The structure and corresponding values used in the BK attacks are listed in Table 4.

TABLE 4. Structure of BK used in breaching privacy.

Known QIDs from external sources	Target SA/QID value
Age, work hours, race, country: [20, $\geq$ 40, white, US]	P(income $\leq$ 50 K)
Age range, profession, marital status: [40-50, Professor, Married]	P(income $>$ 50 K)
Marital status, gender, age, profession: [Widowed, Female, 50, Sales]	P(income $\leq$ 50 K)
Gender, country, income: [Male, China, $>$ 50 K]	P(age $\leq$ 50 Yrs.)
Qualification, race, gender, work class: [Ph.D., White, Female, Private]	P(income $>$ 50 K)

In a linking attack, we computed correct matches based on QID values by amalgamating T and T_s and computed D_{risk}. For fair assessment, we used T_s in different viewpoints, i.e., we considered T_s as BK as well as mainstream data in experimental evaluation. In Figure 9, we present a comparative analysis of privacy results from both attacks (e.g., linking attack and BK attack). For the BK attack, we used a pre-defined structure (given in Table 4), computed D_{risk} from

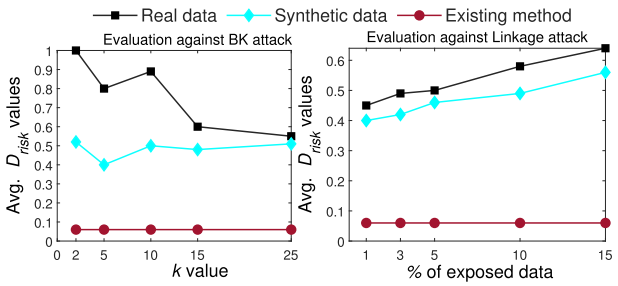


FIGURE 9. Privacy breaches under two distinct attacks.

both T and T_s, and subsequently compared it with the previous study. In the linking attack, we first assume that certain records from both datasets are exposed to the adversary, and then compute D_{risk} value. From the results given in 9, we can see that D_{risk} from T_s was higher than the previously reported generic analysis based on demographics of T_s [58]. In the previous studies, it is often assumed that T_s is not very close to T or generative models do not overfit during training, and therefore, the D_{risk} can be low. However, through experimental analysis, we found that if T_s is of good quality, then privacy attacks can occur with a much higher probability than previously assumed. Through the experiments, we found that D_{risk} can change based on the granularity of information available to the adversary and the amount of data already exposed to public domains. Interestingly, in some cases, we found that D_{risk} between T and T_s is very small, and it differed by just 4% between them. See Case 5 (e.g., when 5% of data is exposed) in Figure 9 (left). The linkage attack on T_s was also possible, and its value is much higher than the previously reported value in the literature. See results in Figure 9 (right) against different %age of exposed data. With the help of this experimental analysis, we determined that the probability of re-identification from T_s can be close to real data, and T_s can also accelerate the privacy attacking process when used improperly (e.g., as BK). These findings further verify our conclusions about AI's threat to privacy.

The values given in Figure 9 were obtained through experiments by using two attack scenarios (BK and linking attacks) that can be applied to anonymized data. We applied these attacks to different anonymized versions produced from real data and synthetic data, respectively. We compared the experimental results with the related SOTA method [58]. From the results, it can be seen that SD curated with our method has yielded higher D_{risk} than the previous method. Furthermore, the D_{risk} values from our method are close to real data in most cases. Based on the above analysis, it is fair to say that AI-generated data can pose a serious threat to an individual's privacy.

D. DATA RECONSTRUCTION ATTACK

In this analysis, we derived various privacy-sensitive rules and analyzed the percentage of data that can be successfully reconstructed utilizing anonymized data and T_s. The QIDs and SAs are part of each rule. Based on the experiments, we found that T_s provides enough knowledge in terms of unique values and their distributions to assist an adversary in converting anonymized data into real data with sufficient accuracy. Subsequently, many attacks can be launched to infer the QIDs/SAs of target individuals from the reconstructed data. We present experimental results from data reconstruction via T_s in Table 5. From the results given in Table 5, a substantial # of tuples can be correctly reconstructed from published data using T_s as BK. We believe this situation (data reconstruction attacks) can be very dangerous when the dataset has skewed distributions (e.g., one value makes the 90% of the data). This scenario and corresponding results highlight the pitfall of AI in the data-sharing scenario.

TABLE 5. Privacy-rules-based data reconstruction via T_s.

Privacy-sensitive rules	Diff (in tuples)	Data re-construction (%)
Unite States $\rightarrow >$ 50 K	2,912	12.33
White $\rightarrow \leq$ 50 K	6,355	27.24
(Doctorate, China) $\rightarrow >$ 50 K	340	5.99
(Male, >40 Hrs. per week) $\rightarrow >>$ 50 K	1,088	46.42
(Exec-Mng, US, White) $\rightarrow >$ 50 K	998	43.60
(>50 Yrs., Divorced, Black)$\rightarrow \leq$ 50 K	328	2.06

The values listed in Table 5 were obtained through matching between anonymized data and T_s using privacy-sensitive rules. Initially, we devised certain privacy rules using the information of QIDs and SA present in the data. Later, we determined the correct matching from the anonymized data against each rule. Since the data is in anonymized form, and therefore, one cannot easily figure out the true values of individuals/users. To reconstruct data, we applied similar rules to synthetic data and chose the relevant records. Afterward, we amalgamated the records from anonymized and T_s to recover the real data. With the help of this method, we were able to reconstruct the real data about users with sufficient accuracy. Through these results, we intend to highlight that T_s can be used as BK, and can assist adversaries in reconstructing the real data from anonymized data.

E. SA PREDICTION ATTACK

In most real-world scenarios, T can be highly skewed, meaning many classes/clusters lack diversity w.r.t SA values. In such cases, if a privacy model such as k-anonymity is applied, explicit privacy disclosures can inevitably occur. Similarly, other models (ℓ-diversity and t-closeness) cannot be directly applied owing to less heterogeneity in SA values. In such circumstances, the adversary can exploit QID values in least-diversity and no-diversity classes and can predict an unknown/new community/individual SA. In Table 6, we present an experiment-based analysis of imbalanced classes in a real-world dataset and the corresponding SA prediction percentage from leveraging T_s. The results show that given the imbalanced information, SAs can be predicted with sufficient accuracy, leading to explicit privacy breaches.

As shown in Table 2, the adult's dataset has a very high imbalance in SA values, and therefore, a significant portion

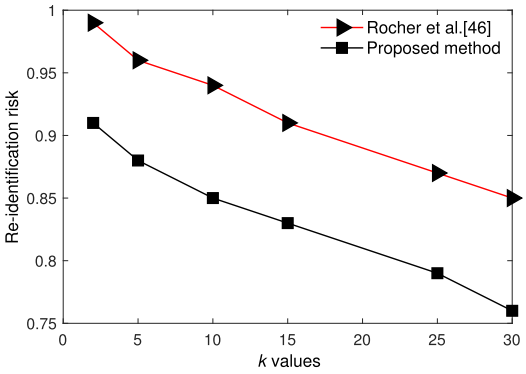
TABLE 6. SA predictions from imbalanced classes using T_s.

k value	# of 0-diverse classes	SA prediction (%)
2	5,000	76.80
4	2,500	70.71
6	900	67.61
8	764	59.87
10	602	56.76
12	543	51.12
14	488	47.03
16	409	40.23
18	392	35.09
20	242	30.71

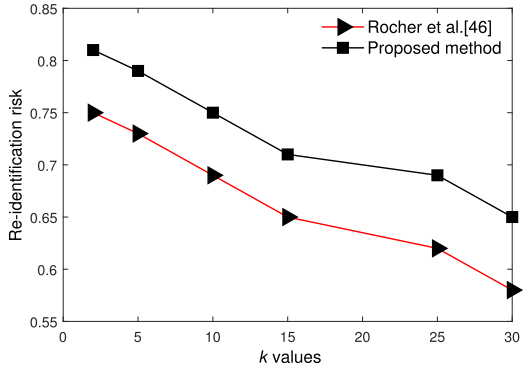
of the data will have no diversity in SA values. If an adversary can somehow link a target user correctly, he/she can infer the SA of that individual, owing to no diversity in the respective class w.r.t. SA. The values in Table 6 were obtained through analysis of zero-diverse classes created with k anonymity model, and using BK (non AI + AI). For example, if an adversary wants to infer the SA of some target user that is located in a non-diverse class, then there is a chance of 100% disclosure. We curated BK from T and T_s and applied it to 0-diverse classes in anonymized data to predict the SA of the target users. From the results and analysis, we found that SA prediction % is high when k is small, owing to a large number of classes with no diversity. Through these results, we intend to highlight that some real-world datasets can be messy, noisy, and imbalanced, which can leak private information when T_s and non-AI BK are jointly used.

F. RE-IDENTICATION ATTACK

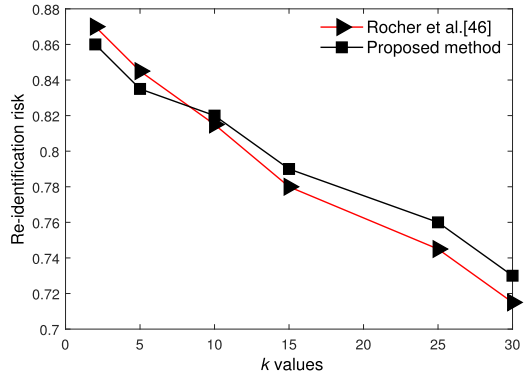
To further validate the effectiveness of our proposal, we computed and compared the re-identification risk (a.k.a. unique re-identification) of an individual from the datasets. Specifically, we compared the success of an individual's re-identification with the method devised by Rocher et al. [59]. In [59], the authors have used generative models to complete the missing values and then reidentify the individuals with very high accuracy. The proposed method can predict the uniqueness of an individual from the dataset with $\geq$ 0.84 AUC. Through fair analysis of the adult dataset, we found that there are about 18, 680 incomplete records (e.g., records with the missing values) in it [14], and the method devised by Rocher et al. [59] has higher re-identification in 40.29% portion of data. In the remaining portion, there is a lower risk of re-identification owing to complete records and the existence of general patterns [39]. For comparison and analysis, we created three partitions (incomplete records partition, privacy-violating partition (records with under-representative value are grouped), and non-privacy-violating partition (records with majorly occurring values were grouped, i.e., in country column, USA value has frequency of $\geq$ 80%, which is regarded as a general pattern and pose less risk to someone's privacy)) of the adult dataset and evaluated the risk of the individual's re-identification. The experimental results and their comparisons are given in Figure 10. From the results, it can be seen that the proposed method yielded a higher re-identification risk than the Rocher et al. [59] method owing to high quality T_s curation. However, in the incomplete records partition, Rocher et al. [59] method has shown better performance owing to incomplete records imputation using GAN models. These results confirm the validity of our proposal and validate the fact that AI can pose threats to individual privacy in data-sharing scenarios.



(a) Incomplete records portion.



(b) Non-privacy voilating portion.



(c) Privacy voilating portion.

FIGURE 10. Re-identification risk: proposed method versus existing method.

The values given in Figure 10 were obtained through experiments by using SD produced with our method and Rocher et al. [59] method. In [59], the authors imputed missing values and computed the re-identification risk from augmented data. We designed three settings of experiments for comparison purposes. As shown in Figure 10, the re-identification risk decreases when k increases in all three cases. In Figure 10(a), the Rocher et al. [59] method has yielded better performance than our method. In the last Figure 10(b), (c), our method has yielded better performance than Rocher et al. [59] method in most cases. The main reason for better performance from our method is due to the higher diversity in SD. Based on the six types of analysis given above, it is fair to say that SD encompasses valuable knowledge that can be used to compromise the privacy of individuals.

To the best of the authors' knowledge, most of the results reported in this paper are new and have not been reported by any of the previous methods thus far. However, in some cases (e.g., subsection V-C,V-F), the SOTA methods were available, and therefore, we fairly compared our results with them. In this paper, we used publicly available datasets (e.g., adults) to verify the effectiveness of our method. However, in real scenarios, the original dataset is not publicly available. Therefore, the extent of unavailable private information that can be retrieved using the AI attack (e.g., SD) can be large subject to the quality of T_s. In practical scenarios, the data owners (e.g., clinics, banks, insurance companies, etc.) do not share/open the data in its original form but allow analysts to generate synthetic copies of data by bringing generative models close to the data. Furthermore, they also share the data in an anonymized form (e.g., removing directly identifiable information and generalizing values of QIDs). Hence, four items can be available to the adversary to compromise individual privacy: (i) AI-based BK (T_s), (ii) non-AI BK (e.g., data/information gathered from other sources, i.e., voter list, factual information, social networking sites, etc.), (iii) anonymization method's understanding (workflow and main steps), and (iv) anonymized data. The previous research has an assumption that SD poses a minimal risk to privacy [60]. However, this work negates that assumption and verifies that privacy risk from SD can be high owing to the quality of T_s. If the quality is good, privacy risk can be high, and vice versa.

If we assume a realistic scenario (e.g., an adult dataset in pure form is unavailable), we were able to deduce useful private information of various kinds that can be used to compromise individual privacy. For example, from the SD and anonymized data, we were able to find that most of the individuals encompassed in T had age values $\geq$ 40 Yrs. Also, the range of age values is between 17 and 90. Similarly, the cardinality information of most attributes was retrieved correctly and can be combined with non-AI BK to compromise someone's privacy. The major pattern (e.g., the earning of people living in the US is over 50 K in most cases) and minor pattern (e.g., the occurrence of records having race value 'other' is very low in the T) in the private information column were retrieved correctly using the AI attack. The attributes leading to group formation, and corresponding group privacy breaches were deduced correctly. The dominant and less dominant values of each attribute were identified correctly using the T_s which can foster identity and SA disclosure. We were able to correctly figure out the SA value and their categories. The indexes of some records were identical in both T_s and T which can enable the linking of target individuals in anonymized data, leading to explicit privacy disclosures.

G. KEY FINDINGS AND MAIN CONCLUSIONS

In this paper, we demonstrate that generative AI models are capable of curating high-quality T_s that constitute proper information of the T, and such data can be used as BK to compromise an individual's privacy in PPDP. To the best of the authors' knowledge, most privacy-preserving mechanisms consider a certain amount of BK that can be available to the adversary, but that is mainly non-AI-based fixed knowledge (e.g., voter lists, online repositories, social network accounts, etc.). To this end, this work highlights another source of BK that has not been properly taken up by the privacy and database communities. However, it needs the attention of the research community because high-quality data can offer helpful hints (e.g., distribution of values, rare and frequent values, patterns, etc.) to the adversary, which can lead to privacy breaches, as demonstrated above. It is worth noting that some works in literature have also highlighted the dangers of AI to information privacy and democracy [61], [62]. We affirm the significance and contributions of the above-cited works, however, these works only highlight the regulatory concerns of AI to information privacy and democracy for legislative actions. Some studies have also explored the closely related aspects such as re-identification risks posed by the generative models to individual privacy [59]. However, these studies have explored only limited risks of AI to information privacy. Our work has enhanced the findings of two closely related existing approaches [58], [59]. Specifically, we highlight the scope of privacy threats that can be launched using AI from a broader perspective and highlight the long-term implications of AI-IP synergy.

The potential ethical implications[4] of using AI as a threat tool to compromise privacy are: downgrading the trust/reputation of data owners, destroying the individualism and self-autonomy of data providers, racism against certain ethnic groups, discrimination against the vulnerable community, illegal intrusion in someone's life, political interest, business sentiment manipulation, illegitimate profiling of the targeted individuals, sensitive rules extraction about different communities, clustering individuals based on controversial behaviors, unequal resource distribution in community beneficial services (e.g., healthcare), lack of trust in government

[4]https://news.harvard.edu/gazette/story/2020/10/ethical-concerns-mount-as-ai-takes-bigger-decision-making-role/

policies, etc. All of the above-cited ethical implications can lead to data silos and the negative use of digital technologies. To address the above challenges, there is an urgent need to establish policies and procedures concerning AI governance. For example, the UNESCO recommendation[5] on ethical AI is one such initiative. In addition, data classification based on sensitivity, and the application of strict privacy-enhancing technologies is another important step to addressing potential ethical implications in data-sharing scenarios. Lastly, integrating AI methods with privacy-enhancing technologies is expected to lower the privacy risks that can stem after data release (when AI acts as a threat tool).

VI. CONCLUDING REMARKS AND IMPLICATIONS

In this paper, we discussed AI synergy in the information-privacy domain from two different aspects (defense and attack). Specifically, we highlight how SD (created with AI tools) can become background knowledge and can assist adversaries in compromising privacy in various ways. We conducted experiments with a real-life benchmark dataset using the CTGAN AI method to prove the feasibility of our idea in real-life scenarios. The experimental results and analysis proved that SD when carefully crafted, can pose various kinds of risks to individual privacy. Our findings showed that SD is another BK source that can enhance the scale and scope of privacy breaches. The new findings offered through this paper are the identification of a new type of BK (e.g., SD) which remained unexplored and underrated in the privacy and database community [32], the negation of the assumption that privacy issues are negligible from SD [60], and the experimental evaluation of privacy risk from a much broader perspective (six different types) that can occur through the amalgamation of AI and non-AI BK in data outsourcing scenarios. To the best of the authors' knowledge, most of the previous research focused on re-identification risk and SA disclosure only, and advanced privacy risks from SD such as group privacy disclosure, data reconstruction, disclosure of SA distributions, and SA prediction from imbalanced classes remained unexplored. Furthermore, our work focuses on breaching the privacy of individuals using AI in PPDP scenarios which is a relatively less investigated topic and has attracted researchers' attention recently. This work draws attention to the invisible risks of AI in the PPDP, which can open up new research dimensions in this line of work. In the future, we intend to develop a practical anonymization mechanism against AI-powered attacks in data-sharing scenarios. Specifically, we aim to propose a threat model by including SD as a new source of BK to safeguard personal data privacy against it. Our algorithm is expected to shed light on AI-powered threats to information privacy which will become a major threat to information privacy in the coming years.

[5]https://www.unesco.org/en/artificial-intelligence/recommendation-ethics/cases

Implications of the proposed work: Although most anonymization methods often assume a certain amount of BK that can be available to the adversary, however, they often assume non-AI knowledge such as voter lists, newspapers, factual information, online repositories, multiple accounts on social networks, etc. Recently, the proliferation of generative tools has enabled the creation of SD that is very close to the real data and can pose serious threats to the privacy of individuals. In Figure 11, we demonstrate a scenario in which SD (e.g., AI-generated knowledge) constitutes BK, and can be leveraged to compromise the privacy of an individual by correctly re-identifying him/her from the published data.

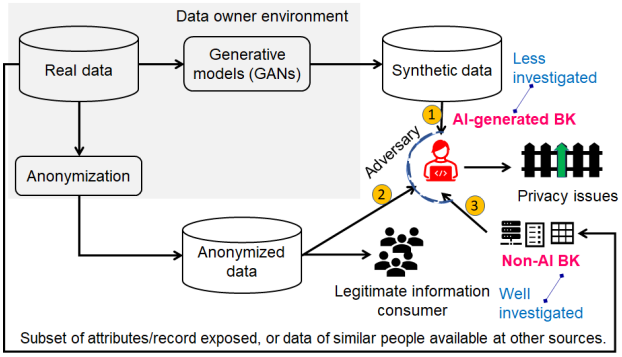


FIGURE 11. Execution of the privacy attacks using SD (e.g., AI-generated BK) combined with non-AI-based BK in a data publishing scenario.

In the past, adversaries usually relied on non-AI knowledge only, but AI-generated knowledge can also be combined with non-AI knowledge to perform identity, sensitive information, and/or membership inference attacks on anonymized data. Therefore, it is vital to realize this changing landscape of privacy breaches amid the rapid proliferation of generative tools and to devise secure privacy-preserving models accordingly [63]. Unfortunately, the privacy implications of SD are relatively unexplored in the privacy and database communities, as they mostly regard SD as an imprecise form of real data. However, this assumption rarely holds, and SD can mimic the properties of real data well in some cases, leading to privacy breaches from anonymized data.

Since SD is mostly generated with AI models, and therefore, AI techniques can be amalgamated with traditional anonymization methods to improve defense against such SD-based attacks. For example, AI techniques can be used to identify attributes that can assist adversaries in compromising someone's identity or sensitive information, hence, improving the defense level. On the other hand, AI techniques can assist in identifying the attributes that have common patterns, and no longer pose threats to the individual's privacy in data sharing [39]. The identification of attributes that are not vulnerable to privacy can be minimally generalized, leading to higher truthfulness in data. To this end, our results and findings can guide the privacy and database communities to integrate AI techniques with traditional anonymization

methods to the extent possible to improve the performance bottlenecks of the existing anonymization models.

In recent years, privacy mechanisms have been increasingly used with AI models (e.g., federated learning) to protect the privacy of training data, parameters, and data memorization issues [64], [65], [66]. Since privacy mechanisms and AI techniques have become indispensable components of each other, and therefore, this work offers important directions to investigate the relationship between these two concepts that are not directly related. Lastly, this work presents helpful knowledge about the invisible risks of AI in data-sharing scenarios that can assist data owners in considering AI-powered BK as well while outsourcing the data, leading to better data governance and use in the AI-driven era.

CONFLICT OF INTEREST

The authors declare that they have no conflict of interest.

ACKNOWLEDGMENT

The authors sincerely thank the associate editor and six expert reviewers who thoroughly evaluated this article and provided very constructive feedback, which significantly enhanced the quality of this article.

REFERENCES

[1] N. L. Bragazzi, H. Dai, G. Damiani, M. Behzadifar, M. Martini, and J. Wu, "How big data and artificial intelligence can help better manage the COVID-19 pandemic," *Int. J. Environ. Res. Public Health*, vol. 17, no. 9, p. 3176, May 2020, doi: 10.3390/ijerph17093176.

[2] B. C. Kara and C. Eyupoglu, "Anonymization methods for privacy-preserving data publishing," in *Smart Applications With Advanced Machine Learning and Human-Centred Problem Design*. Cham, Switzerland: Springer, 2023, pp. 145–159.

[3] F. Liang, F. Liu, and T. Zhou, "A visual tool for interactively privacy analysis and preservation on order-dynamic tabular data," in *Collaborative Computing: Networking, Applications and Worksharing*. Cham, Switzerland: Springer, 2023, pp. 18–38.

[4] C. Dhasarathan, M. K. Hasan, S. Islam, S. Abdullah, U. A. Mokhtar, A. R. Javed, and S. Goundar, "COVID-19 health data analysis and personal data preserving: A homomorphic privacy enforcement approach," *Comput. Commun.*, vol. 199, pp. 87–97, Feb. 2023.

[5] M. Hernadez, G. Epelde, A. Alberdi, R. Cilla, and D. Rankin, "Synthetic tabular data evaluation in the health domain covering resemblance, utility, and privacy dimensions," *Methods Inf. Med.*, vol. 62, no. S01, pp. e19–e38, Jun. 2023.

[6] L. Sweeney, "K-anonymity: A model for protecting privacy," *Int. J. Uncertainty, Fuzziness Knowl.-Based Syst.*, vol. 10, no. 5, pp. 557–570, Oct. 2002.

[7] A. Machanavajjhala, D. Kifer, J. Gehrke, and M. Venkitasubramaniam, "ℓ-diversity: Privacy beyond K-anonymity," *ACM Trans. Knowl. Discovery Data*, vol. 1, no. 1, pp. 1–12, 2007.

[8] N. Li, T. Li, and S. Venkatasubramanian, "T-closeness: Privacy beyond K-anonymity and L-diversity," in *Proc. IEEE 23rd Int. Conf. Data Eng.*, Apr. 2007, pp. 106–115.

[9] C. Dwork, "Differential privacy: A survey of results," in *Proc. Int. Conf. Theory Appl. Models Comput.* Cham, Switzerland: Springer, 2008, pp. 1–19.

[10] J. Jia, C. Tan, Z. Liu, X. Li, Z. Liu, S. Lv, and C. Dong, "Total variation distance privacy: Accurately measuring inference attacks and improving utility," *Inf. Sci.*, vol. 626, pp. 537–558, May 2023, doi: 10.1016/j.ins.2023.01.037.

[11] S. Srijayanthi and T. Sethukarasi, "Design of privacy preserving model based on clustering involved anonymization along with feature selection," *Comput. Secur.*, vol. 126, Mar. 2023, Art. no. 103027, doi: 10.1016/j.cose.2022.103027.

[12] A. Majeed and S. O. Hwang, "Quantifying the vulnerability of attributes for effective privacy preservation using machine learning," *IEEE Access*, vol. 11, pp. 4400–4411, 2023, doi: 10.1109/ACCESS.2023.3235016.

[13] N. Jha, L. Vassio, M. Trevisan, E. Leonardi, and M. Mellia, "Practical anonymization for data streams: Z-anonymity and relation with K-anonymity," *Perform. Eval.*, vol. 159, Jan. 2023, Art. no. 102329, doi: 10.1016/j.peva.2022.102329.

[14] L. Chen, L. Zeng, Y. Mu, and L. Chen, "Global combination and clustering based differential privacy mixed data publishing," *IEEE Trans. Knowl. Data Eng.*, early access, Jan. 17, 2023, doi: 10.1109/TKDE.2023.3237822.

[15] B. Su, J. Huang, K. Miao, Z. Wang, X. Zhang, and Y. Chen, "K-anonymity privacy protection algorithm for multi-dimensional data against skewness and similarity attacks," *Sensors*, vol. 23, no. 3, p. 1554, Jan. 2023, doi: 10.3390/s23031554.

[16] Y. S. Hindistan and E. F. Yetkin, "A hybrid approach with GAN and DP for privacy preservation of IIoT data," *IEEE Access*, vol. 11, pp. 5837–5849, 2023, doi: 10.1109/ACCESS.2023.3235969.

[17] P. Tang, R. Chen, S. Su, S. Guo, L. Ju, and G. Liu, "Multi-party sequential data publishing under differential privacy," *IEEE Trans. Knowl. Data Eng.*, early access, Feb. 2, 2023, doi: 10.1109/TKDE.2023.3241661.

[18] M. Orooji, S. S. Rabbanian, and G. M. Knapp, "Flexible adversary disclosure risk measure for identity and attribute disclosure attacks," *Int. J. Inf. Secur.*, vol. 22, pp. 1–15, Jan. 2023.

[19] Y. Wang, Y. Luo, L. Liu, and S. Fu, "pCOVID: A privacy-preserving COVID-19 inference framework," in *Algorithms and Architectures for Parallel Processing*. Copenhagen, Denmark: Springer, 2023, pp. 21–42.

[20] A. Soliman, S. Rajasekaran, P. Toman, and N. Ravishanker, "A fast privacy-preserving patient record linkage of time series data," *Sci. Rep.*, vol. 13, no. 1, pp. 1–10, Feb. 2023.

[21] M. Sundaram, M. S. Annamalai, A. Gadotti, and L. Rocher, "A linear reconstruction approach for attribute inference attacks against synthetic data," 2023, *arXiv:2301.10053*.

[22] M. Hittmeir, A. Ekelhart, and R. Mayer, "Utility and privacy assessments of synthetic data for regression tasks," in *Proc. IEEE Int. Conf. Big Data (Big Data)*, Dec. 2019, pp. 5763–5772.

[23] C. Little, M. Elliot, and R. Allmendinger, "Comparing the utility and disclosure risk of synthetic data with samples of microdata," in *Proc. Privacy Stat. Databases, Int. Conf. (PSD)*. Cham, Switzerland: Springer, 2022, pp. 234–249. [Online]. Available: https://crises-deim.urv.cat/psd2022/

[24] N. Ruiz, K. Muralidhar, and J. Domingo-Ferrer, "On the privacy guarantees of synthetic data: A reassessment from the maximum-knowledge attacker perspective," in *Privacy in Statistical Databases: UNESCO Chair in Data Privacy*. Cham, Switzerland: Springer, 2018, pp. 59–74.

[25] M. Hittmeir, R. Mayer, and A. Ekelhart, "A baseline for attribute disclosure risk in synthetic data," in *Proc. 10th ACM Conf. Data Appl. Secur. Privacy*, Mar. 2020, pp. 133–143.

[26] T. Yang, L. S. Cang, M. Iqbal, and D. Almakhles, "Attack risk analysis in data anonymization in Internet of Things," *IEEE Trans. Computat. Social Syst.*, early access, Feb. 22, 2023, doi: 10.1109/TCSS.2023.3243089.

[27] B. Liu, M. Ding, S. Shaham, W. Rahayu, F. Farokhi, and Z. Lin, "When machine learning meets privacy: A survey and outlook," *ACM Comput. Surveys*, vol. 54, no. 2, pp. 1–36, Mar. 2022.

[28] A. Majeed and S. O. Hwang, "A practical anonymization approach for imbalanced datasets," *IT Prof.*, vol. 24, no. 1, pp. 63–69, Jan. 2022.

[29] P. Silva, C. Gonçalves, N. Antunes, M. Curado, and B. Walek, "Privacy risk assessment and privacy-preserving data monitoring," *Exp. Syst. Appl.*, vol. 200, Aug. 2022, Art. no. 116867, doi: 10.1016/j.eswa.2022.116867.

[30] M. M. H. Onik, C.-S. Kim, and J. Yang, "Personal data privacy challenges of the fourth industrial revolution," in *Proc. 21st Int. Conf. Adv. Commun. Technol. (ICACT)*, Feb. 2019, pp. 635–638.

[31] X. Ding, H. Zhang, C. Ma, X. Zhang, and K. Zhong, "User identification across multiple social networks based on naive Bayes model," *IEEE Trans. Neural Netw. Learn. Syst.*, early access, Sep. 14, 2022, doi: 10.1109/TNNLS.2022.3202709.

[32] N. Desai, M. Lal Das, P. Chaudhari, and N. Kumar, "Background knowledge attacks in privacy-preserving data publishing models," *Comput. Secur.*, vol. 122, Nov. 2022, Art. no. 102874, doi: 10.1016/j.cose.2022.102874.

[33] L. Xu, M. Skoularidou, A. Cuesta-Infante, and K. Veeramachaneni, "Modeling tabular data using conditional GAN," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 32, 2019, pp. 1–15.

[34] A. Majeed and S. O. Hwang, "Rectification of syntactic and semantic privacy mechanisms," *IEEE Secur. Privacy*, early access, Jul. 29, 2022, doi: 10.1109/MSEC.2022.3188365.
[35] J. Jayaram, P. Manickam, A. Gupta, and M. Rudrabhatla, "CBPP: An efficient algorithm for privacy-preserving data publishing of 1: M micro data with multiple sensitive attributes," in *Machine Learning Algorithms and Applications in Engineering*. Boca Raton, FL, USA: CRC Press, 2023, p. 195, doi: 10.1201/9781003104858.
[36] C. Ni, L. S. Cang, P. Gope, and G. Min, "Data anonymization evaluation for big data and IoT environment," *Inf. Sci.*, vol. 605, pp. 381–392, Aug. 2022.
[37] C. Zhang, H. Jiang, Y. Wang, Q. Hu, J. Yu, and X. Cheng, "User identity de-anonymization based on attributes," in *Wireless Algorithms, Systems, and Applications*. Cham, Switzerland: Springer, 2019, pp. 458–469.
[38] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.
[39] M. Strobel and R. Shokri, "Data privacy and trustworthy machine learning," *IEEE Secur. Privacy*, vol. 20, no. 5, pp. 44–49, Sep. 2022.
[40] T. Kim and J. Yang, "Selective feature anonymization for privacy-preserving image data publishing," *Electronics*, vol. 9, no. 5, p. 874, May 2020, doi: 10.3390/electronics9050874.
[41] M. Maximov, I. Elezi, and L. Leal-Taixé, "CIAGAN: Conditional identity anonymization generative adversarial networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 5446–5455.
[42] D. N. Jaidan, M. Carrere, Z. Chemli, and R. Poisvert, "Data anonymization for privacy aware machine learning," in *Machine Learning, Optimization, and Data Science*. Cham, Switzerland: Springer, 2019, pp. 725–737.
[43] N. Senavirathne and V. Torra, "On the role of data anonymization in machine learning privacy," in *Proc. IEEE 19th Int. Conf. Trust, Secur. Privacy Comput. Commun. (TrustCom)*, Dec. 2020, pp. 664–675.
[44] S. Shaham, M. Ding, B. Liu, S. Dang, Z. Lin, and J. Li, "Privacy preserving location data publishing: A machine learning approach," *IEEE Trans. Knowl. Data Eng.*, vol. 33, no. 9, pp. 3270–3283, Sep. 2021.
[45] R. Bryant, C. Cintas, I. Wambugu, A. Kinai, A. Diriye, and K. Weldemariam, "Evaluation of bias in sensitive personal information used to train financial models," in *Proc. IEEE Global Conf. Signal Inf. Process. (GlobalSIP)*, Nov. 2019, pp. 1–5.
[46] J. Tomás, D. Rasteiro, and J. Bernardino, "Data anonymization: An experimental evaluation using open-source tools," *Future Internet*, vol. 14, no. 6, p. 167, May 2022, doi: 10.3390/fi14060167.
[47] D. Jeong, J. H. T. Kim, and J. Im, "A new global measure to simultaneously evaluate data utility and privacy risk," *IEEE Trans. Inf. Forensics Security*, vol. 18, pp. 715–729, 2023.
[48] T. Zhu, D. Ye, W. Wang, W. Zhou, and P. S. Yu, "More than privacy: Applying differential privacy in key areas of artificial intelligence," *IEEE Trans. Knowl. Data Eng.*, vol. 34, no. 6, pp. 2824–2843, Jun. 2022.
[49] D. Ye, S. Shen, T. Zhu, B. Liu, and W. Zhou, "One parameter defense—Defending against data inference attacks via differential privacy," *IEEE Trans. Inf. Forensics Security*, vol. 17, pp. 1466–1480, 2022, doi: 10.1109/TIFS.2022.3163591.
[50] L. Hu, J. Li, G. Lin, S. Peng, Z. Zhang, Y. Zhang, and C. Dong, "Defending against membership inference attacks with high utility by GAN," *IEEE Trans. Dependable Secure Comput.*, vol. 20, no. 3, pp. 2144–2157, Jun. 2022, doi: 10.1109/TDSC.2022.3174569.
[51] H. Liu, D. Xu, Y. Tian, C. Peng, Z. Wu, and Z. Wang, "Wasserstein generative adversarial networks based differential privacy metaverse data sharing," *IEEE J. Biomed. Health Informat.*, early access, Jun. 16, 2023, doi: 10.1109/JBHI.2023.3287092.
[52] E. Choi, S. Biswal, B. Malin, J. Duke, W. F. Stewart, and J. Sun, "Generating multi-label discrete patient records using generative adversarial networks," in *Proc. Mach. Learn. Healthcare Conf.*, 2017, pp. 286–305.
[53] Y. Zhang, N. Zaidi, J. Zhou, and G. Li, "Interpretable tabular data generation," *Knowl. Inf. Syst.*, vol. 65, pp. 1–29, Jan. 2023.
[54] C. M. Bishop and N. M. Nasrabadi, *Pattern Recognition and Machine Learning*, vol. 4, no. 4. Cham, Switzerland: Springer, 2006. [Online]. Available: https://link.springer.com/book/9780387310732
[55] Z. Lin, A. Khetan, G. Fanti, and S. Oh, "PacGAN: The power of two samples in generative adversarial networks," *IEEE J. Sel. Areas Inf. Theory*, vol. 1, no. 1, pp. 324–335, May 2020.
[56] H. P. Das, R. Tran, J. Singh, X. Yue, G. Tison, A. Sangiovanni-Vincentelli, and C. J. Spanos, "Conditional synthetic data generation for robust machine learning applications with limited pandemic data," in *Proc. AAAI Conf. Artif. Intell.*, vol. 36, no. 11, 2022, pp. 11792–11800.
[57] K. Fang, V. Mugunthan, V. Ramkumar, and L. Kagal, "Overcoming challenges of synthetic data generation," in *Proc. IEEE Int. Conf. Big Data (Big Data)*, Dec. 2022, pp. 262–270.
[58] K. El Emam, L. Mosquera, E. Jonker, and H. Sood, "Evaluating the utility of synthetic COVID-19 case data," *JAMIA Open*, vol. 4, no. 1, Mar. 2021, Art. no. ooab012, doi: 10.1093/jamiaopen/ooab012.
[59] L. Rocher, J. M. Hendrickx, and Y.-A. de Montjoye, "Estimating the success of re-identifications in incomplete datasets using generative models," *Nature Commun.*, vol. 10, no. 1, pp. 1–9, Jul. 2019.
[60] K. El Emam, "Seven ways to evaluate the utility of synthetic data," *IEEE Secur. Privacy*, vol. 18, no. 4, pp. 56–59, Jul. 2020.
[61] K. Manheim and L. Kaplan, "Artificial intelligence: Risks to privacy and democracy," *Yale JL Tech.*, vol. 21, no. 1, p. 106, Annual 2019.
[62] M. Sigalas. (2021). *Artificial Intelligence: Dangers to Privacy and Democracy*. [Online]. Available: http://hdl.handle.net/11610/23534
[63] A. Kiran and S. S. Kumar, "Synthetic data and its evaluation metrics for machine learning," in *Information Systems for Intelligent Systems*. Cham, Switzerland: Springer, 2023, pp. 485–494.
[64] A. K. Nair, J. Sahoo, and E. D. Raj, "Privacy preserving federated learning framework for IoMT based big data analysis using edge computing," *Comput. Standards Interfaces*, vol. 86, Aug. 2023, Art. no. 103720, doi: 10.1016/j.csi.2023.103720.
[65] Y. Zhou, X. Liu, Y. Fu, D. Wu, J. H. Wang, and S. Yu, "Optimizing the numbers of queries and replies in convex federated learning with differential privacy," *IEEE Trans. Dependable Secure Comput.*, early access, Jan. 6, 2023, doi: 10.1109/TDSC.2023.3234599.
[66] J. Chen, J. Xue, Y. Wang, L. Huang, T. Baker, and Z. Zhou, "Privacy-preserving and traceable federated learning for data sharing in industrial IoT applications," *Exp. Syst. Appl.*, vol. 213, Mar. 2023, Art. no. 119036, doi: 10.1016/j.eswa.2022.119036.

ABDUL MAJEED received the B.S. degree in information technology from UIIT, PMAS-UAAR, Rawalpindi, Pakistan, in 2013, the M.S. degree in information security from COMSATS University, Islamabad, Pakistan, in 2016, and the Ph.D. degree in computer information systems and networks from Korea Aerospace University, South Korea, in 2021. He was a Security Analyst with Trillium Information Security Systems (TISS), Rawalpindi, from 2015 to 2016. He is currently an Assistant Professor with the Department of Computer Engineering, Gachon University, South Korea. His research interests include privacy-preserving data publishing, statistical disclosure control, privacy-aware analytics, data-centric AI, and machine learning.

SEONG OUN HWANG (Senior Member, IEEE) received the B.S. degree in mathematics from Seoul National University, in 1993, the M.S. degree in information and communications engineering from the Pohang University of Science and Technology, in 1998, and the Ph.D. degree in computer science from the Korea Advanced Institute of Science and Technology, South Korea, in 2004. He was a Software Engineer with LG-CNS Systems Inc., from 1994 to 1996. He was a Senior Researcher with the Electronics and Telecommunications Research Institute (ETRI), from 1998 to 2007, and a Professor with the Department of Software and Communications Engineering, Hongik University, from 2008 to 2019. He is currently a Full Professor with the Department of Computer Engineering, Gachon University, South Korea. His research interests include cryptography, data-centric AI, cybersecurity, and artificial intelligence.