iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 12 **Issue:** V **Month of publication:** May 2024

DOI: https://doi.org/10.22214/ijraset.2024.61560

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue V May 2024- Available at www.ijraset.com

AI & The Privacy Paradox: Navigating Synthetic Data Vaults

Nithin Manoj[1], Pavithran M A[2], Sivani Binu[3], Sreedevi Prasad[4], Asha J George[5]
Department of Computer Science and Engineering Mar Athanasius College of Engineering, Kothamangalam, Kerala

Abstract: ***Exploring the synergies and threats of artificial intelligence in the information privacy domain, the focus is on synthetic data (SD) as a form of background knowledge (BK) that can compromise privacy in data publishing scenarios. A novel taxonomy of BK that includes AI-generated data is proposed, demonstrating how it can lead to various kinds of privacy breaches, such as re-identification, sensitive attribute disclosure, and group privacy disclosure. Experiments are conducted on a real-life benchmark dataset using a state-of-the-art AI model to generate SD, showing that SD can pose serious risks to individual privacy when carefully crafted and used by adversaries. In realtime processing paradigms this method may face some challenges such as the highly customized implementation of libraries needed for computing statistics and visualizations of query answers for robust analysis without jeopardizing user's privacy. To overcome the challenges in real-time processing paradigms, the algorithm can use stream anonymization methods or online learning techniques to anonymize data in an incremental and adaptive manner.***

Keywords: Sensitive Attributes, Background Knowledge, Conditional Tabular Generative Adversarial Network (CTGAN), Quasi Identifiers (QIDs)

I. INTRODUCTION

This project focuses on developing a Synthetic Data Vaults tool for social platforms. The rise of toxic and adversarial attacks on the real data in online communication necessitates the development of robust privacy-preserving models to lower the risk of attribute disclosure and re-identification attacks. The objective is to provide a secure and privacy-preserving environment for storing and managing synthetic data. It allows organizations to generate data that closely resembles real data without exposing sensitive information.

II. RELATED WORKS

A. Threats of Artificial Intelligence in Privacy Preservation

Beyond traditional frameworks, newer studies have placed more emphasis on using artificial intelligence (AI) methods into privacy protection initiatives. Among the tactics investigated in this field are homomorphic encryption, federated learning, and differential privacy. The use of synthetic data (SD) as a background knowledge (BK) element is still largely unexplored, despite these advancements. By putting out a fresh taxonomy that incorporates AI-generated data into the privacy discourse, our work aims to close this gap[6]. By means of this integration, the research illuminates the subtle dangers that SD poses in data publishing situations, enhancing our comprehension of privacy hazards in the age of artificial intelligence-driven data synthesis.

B. Taxonomy of Background Knowledge in Privacy Breaches

In earlier times, conversations about privacy have mostly focused on well-known flaws such as group privacy breaches, sensitive attribute disclosure, and re-identification attacks. Although these are still important issues, a more comprehensive view is required given the way that data collection and utilization are changing. Our taxonomy broadens this discussion by including AI-generated data as a potent type of background information. This taxonomy offers a thorough foundation for comprehending the many hazards associated with using SD in privacy-sensitive situations through organized classification. Crafting effective privacy safeguards that anticipate and handle emerging threats to data privacy requires a comprehensive grasp of the subject.

C. Empirical Analysis of Synthetic Data Risks

There doesn't appear to be significant actual support for the risks associated with the synthetic data in the scientific community, despite theoretical assumptions. This work closes this gap by carrying out careful tests on an actual benchmark dataset. Modern AI models are used to generate and analyze synthetic data in a methodical manner in order to assess its privacy impact.

The results highlight the significant hazards associated with well-designed SD, which emphasizes the need for strong privacy-preserving measures[9]. Moreover, the empirical data provides a solid foundation for well-informed decision-making about the creation of technologically sophisticated privacy safeguards and privacy policies.

D. Challenges and Solutions in Real-Time Processing

Integrating privacy-preserving methods poses significant hurdles in the field of real-time data processing. One major challenge is the highly customized development of libraries for computing statistics and visualizations while protecting user privacy. The use of online learning strategies for incremental and adaptive data anonymization as well as stream anonymization techniques are suggested solutions to these problems. These tactics support the creation of scalable and privacy-respecting data analytics frameworks in dynamic contexts by enabling real-time processing capabilities while preserving individual privacy rights. To further strengthen privacy-preserving frameworks in dynamic data settings, investigating the incorporation of differential privacy techniques within real-time processing systems shows promise in striking a balance between data value and privacy protection.

E. Ethical Considerations and Societal Implications

With AI technologies developing further, data privacy ethics are becoming more and more important. Ensuring the ethical and transparent implementation of privacypreserving measures while taking into account any societal ramifications is of utmost importance. In order to protect fundamental rights and values, ethical frameworks must direct the development and application of AI systems, regardless of their technological effectiveness. This means abiding by moral standards like justice, accountability, and transparency in addition to following the law. Furthermore, ethical considerations become even more crucial as AI systems progressively impact decision-making processes across a variety of fields, including healthcare, finance, and criminal justice. The necessity of taking a proactive stance when it comes to the development and control of ethical AI is highlighted by the possibility of algorithmic bias, prejudice, and unforeseen repercussions.

III. PROPOSED MODEL

Proposed algorithm to mitigate the possible risk of synthetic data breaching an individual's privacy. The method by which the suggested algorithm addresses the possible risk that fake data poses. There are three main modules of the proposed algorithm: 1. Synthetic data generation using generative AI models 2. AI and non-AI BK-aware anonymization of real data 3. Anonymization data sharing.

A. Synthetic Data Generation Using Generative AI Models

SD is produced by imitating actual data's characteristics. Many generative AI techniques have been put out in recent years to either maximize the creation of synthetic data or curate high-quality data. The suggested approach uses CTGAN to generate SD, which is then used as BK to measure the anonymization algorithm's strengths[8],[10] . A feature of the CTGAN model allows for the correct modelling of all values and their distributions.

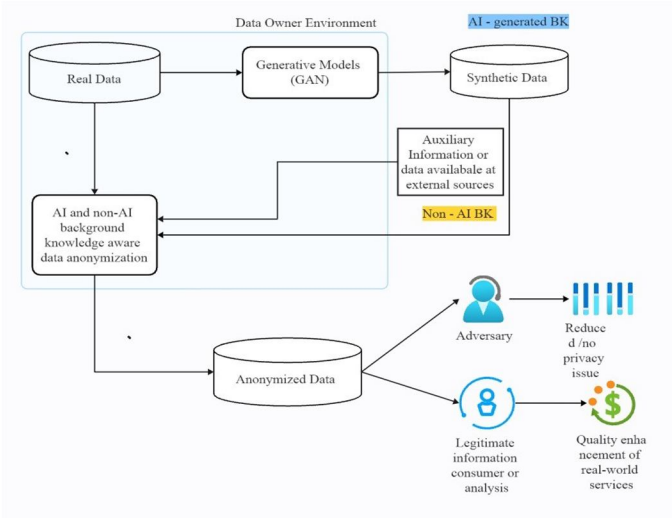


Fig. 1. Proposed Design

B. 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 defence against breaching an individual's privacy. Seven key steps in the proposed algorithm are sequentially applied to generate anonymized data from T. In any real-world T, there are four types of attributes: sensitive attributes (SA), quasiidentifiers (QIDs), non-sensitive attributes (NSAs), and explicit identifiers (EIs)[1].

In first step data is cleaned in the data cleaning step with the aid of advanced pre-processing methods. This step's primary goals are to simplify data interpretation and guard against drawing incorrect inferences from anonymised data. Furthermore, the data is cleansed to avoid any issues with anonymization. Using the information about record similarity, duplicates are eliminated. Records that are next to one another are presumed to be related. Before anonymizing the data, augmentation and distribution analysis can be used to enhance the quality of the data if T is of low quality.

$$T = \begin{array}{c|cccc} x_i\,x & q_1 & q_2 & \ldots & x_p \quad Y \\ 1 & vqx11 & vqx21 & \ldots & vqxp1 \quad y1 \\ x_2 & vqx12 & vqx22 & \ldots & vqxp2 \quad y2 \\ \vdots & \ldots & \ldots & \ldots & \ldots \quad \vdots \\ \ldots & vqx1N & vqx2N & \ldots & vqxpN \\ xN & & & & yp' \end{array}$$

The second step involves computing statistics about QIDs' utility, vulnerability, information gain, etc. We used the random forest ensemble method to compute vulnerability[3]. One of the dependable machine learning techniques with a wide range of applications is random forest. To order the QIDs according to vulnerability, we used the RF technique. We use both original and shuffled data to build the model[2],[7].

In the process of grouping records, we calculate the similarity S between them. Records are grouped throughout the clustering process according to the S value and the k anonymity criterion, which requires that each cluster contain a minimum of k records.

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

Because compact clusters are produced by our technique, anonymization processes result in smaller generalization intervals. Information loss is limited because every cluster has the same records. Uncertainty U is calculated from the SA column following the creation of clusters.

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

The QIDs' initial values are changed to more generalized ones in the following stage. Two different procedures were used to convert QID values: laplace noise addition and generalization hierarchy. Using the laplace method to anonymize numerical data is a great idea. Using laplace techniques, the anonymized numerical QID (q) can be produced as

$$q' = q + n \qquad (5)$$

The anonymized data is then assessed for privacy before being made available to academics and data miners in the following stage. While existing approaches only consider the non-AI BK, the proposed algorithm considers both AI-based and non-AI BK to assess the level of privacy in anonymized datasets[3]. Strict privacy guarantees cannot be achieved in data publishing scenarios by taking only non-AI knowledge into account, and the likelihood of explicit privacy breaches is considerable when the adversary has access to some highquality SD. A number of records as a BK are taken from auxiliary sources and SD during the assessment of privacy strengths, and their privacy disclosure is carried out using anonymized data. In addition, additional valuable insights (such as minority values, prevalent patterns, frequency of values, etc.) obtained from the SD were also utilized to ascertain the privacy status of specific individuals in anonymized data. In order to boost the defensive level, stricter settings for the privacy parameter (such as k, , etc.) were applied if the privacy disclosures in anonymized data were high. After passing all privacy requirements, data can be contracted out for the purpose of knowledge discovery. In certain circumstances, SD can also be used with the actual data to enhance distribution skewness or reduce the number of records, which improves data anonymization.

Finally, parameter tuning for stricter privacy parameter values were used to boost protection levels if there were a lot of privacy disclosures in anonymized data.

After passing all privacy requirements, data can be contracted out for the purpose of knowledge discovery. In certain circumstances, SD can also be used with the actual data to enhance distribution skewness or reduce the number of records, which improves data anonymization.

C. Anonymized Data Sharing for Analytics While Preserving the Privacy of Individuals

This module provides analysts, researchers, and/or data miners with anonymized data so they can draw insights without jeopardizing the privacy of individual users. Since the anonymized data in the suggested algorithm is subjected to stringent privacy checks, there is little chance of privacy violations[4]. Since the suggested algorithm treats SD as BK while anonymizing data, the likelihood of privacy risk in anonymized data can be limited. In contrast to the current approach, which only takes non-AI BK into account and increases privacy breaches, the suggested algorithm can offer a superior safety in the presence of both AI and non-AI BK. Finally, our system retains more knowledge for information consumers by taking use of generalization hierarchy and differential privacy. Our technique preserves superior interpretations of T in T by taking advantage of intrinsic properties of attributes in T, which results in increased utility and privacy. It is important to remember that every anonymization technique can be strengthened to fend against AI-powered attacks by adding one more step (such as taking into account any AI-based BK that the opponent may possess)[5]. Our technique is general-purpose and may be used to offer strong protection in data-sharing scenarios against AI-based and non-AI-based BK.

IV. RESULTS

We conducted thorough experiments to show the adversarial function of AI in data publishing using a real-world dataset. a standard benchmark dataset with a range of sensitivity attribute (SA) values and mixed attribute types (e.g., numerical and categorical) that is used to show the practicality of anonymization algorithms. We also eliminated any duplicate entries that were adjacent to one another and had values that were the same row after row. We were able to accurately measure the amount of privacy leaks by using an error-free dataset for our trials.

Overall Score: 100.0%

Properties:

Data Validity: 100.0%

Data Structure: 100.0%

All primary keys must be unique. Continuous values must adhere to the min/max of the real data, discrete columns (non-PII) must have the same categories as the real data.

Overall Score: 89.53 Properties:

Column Shapes: 91.76

Column Pair Trends: 87.29

According to the score, the synthetic data is about 89 % like the real data in terms of statistical similarity.

Real Data

Synthetic Data

TABLE I
QUALITY OF EACH DATA COLUMN

Column	Metric	Score
has rewards	TVComplement	0.976000
room type	TVComplement	0.914000
amenities fee	KSComplement	0.931136
checkin date	KSComplement	0.934000
checkout date	KSComplement	0.918750
room rate	KSComplement	0.832000

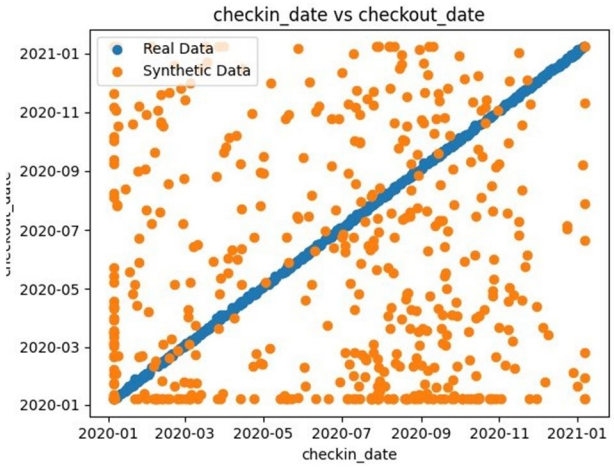


Fig. 2. Column Pair Plot of Attributes 'checkin date' and 'checkout–date'

V. FUTURE SCOPE

The project "AI & the Privacy Paradox: Navigating Synthetic Data Vaults" paves the way for future research and development in artificial intelligence, privacy preservation, and synthetic data management. As online communication evolves, the importance of robust privacy-preserving models and tools becomes increasingly critical.

1) *Adaptation to Real-World Scenarios:* Future iterations of the Synthetic Data Vaults tool can be tailored to adapt to diverse real-world scenarios, spanning various social platforms and communication channels. This adaptability will enable the tool to address specific privacy challenges inherent to different platforms, enhancing its applicability and effectiveness.

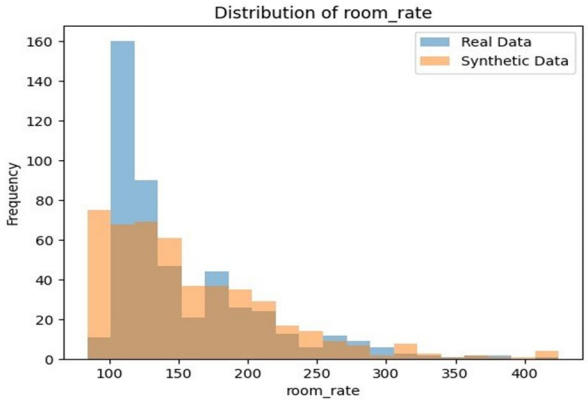


Fig. 3. Column Plot of Frequency Distribution of the Attribute 'room rate'

2) *Integration with Differential Privacy (DP) Models:* Integrating DP models can further strengthen the project's privacy-preserving capabilities by ensuring that data releases do not inadvertently disclose sensitive information about individual users. This integration will bolster user trust and confidence in the system's ability to protect their privacy while maintaining data utility.
3) *Data Augmentation for Machine Learning:* Leveraging synthetic data for data augmentation in machine learning applications offers a promising avenue for enhancing model performance and robustness. This approach can improve model generalization and reduce reliance on realworld data.
4) *Enhanced Anonymization Techniques:* With evolving data privacy threats, there is a growing need to develop advanced anonymization techniques capable of thwarting sophisticated privacy attacks. Future research could explore innovative stream anonymization methods, online learning techniques, and advanced encryption algorithms to enhance the project's data anonymization capabilities.

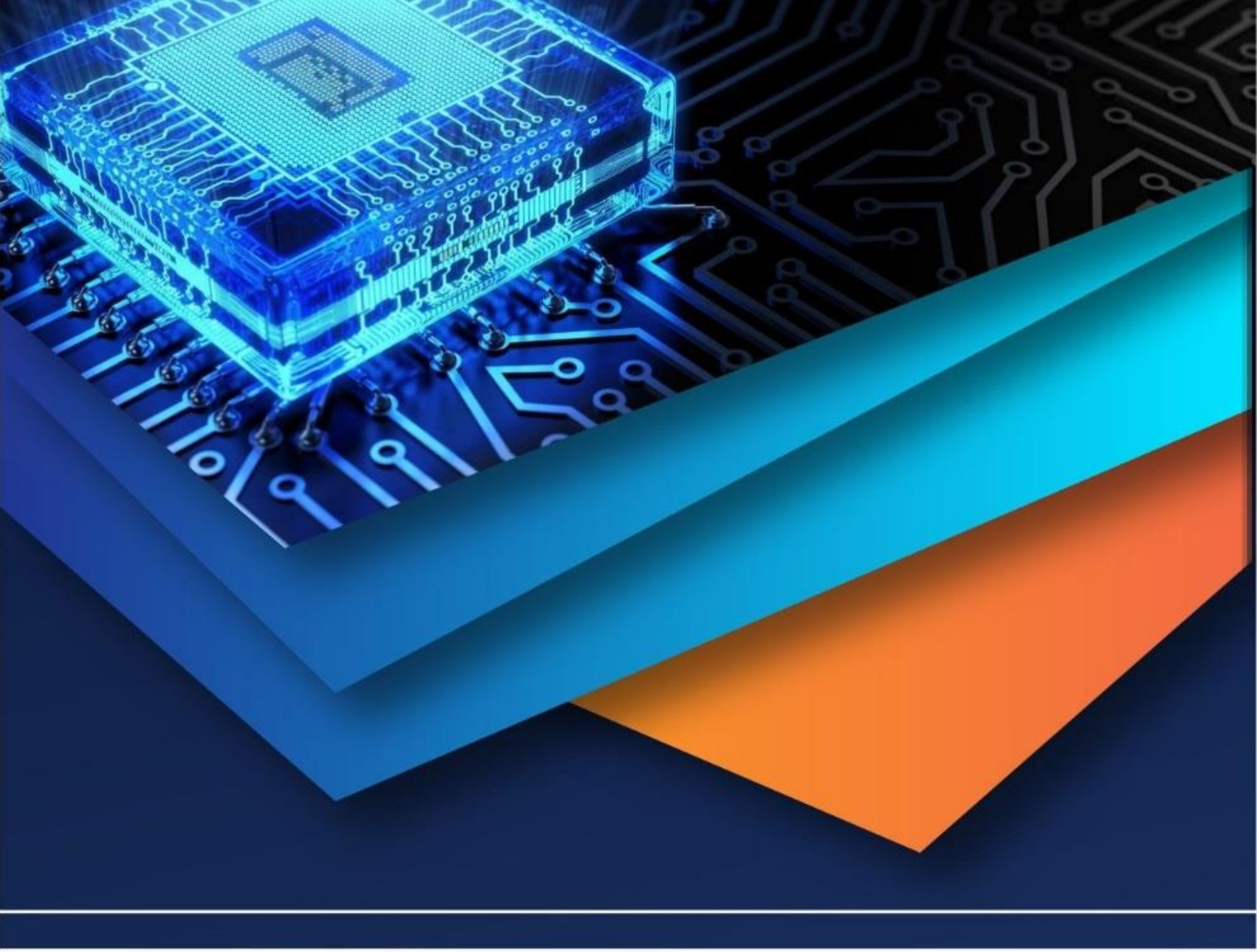
5) *User-Centric Features and Customization:* To ensure widespread adoption and user engagement, future developments could prioritize user-centric features and customization options. This may include personalized privacy settings, interactive dashboards for data visualization, and user-friendly interfaces, empowering users with greater control over their data.

VI. CONCLUSION

The project ”AI & the Privacy Paradox: Navigating Synthetic Data Vaults” underscores the critical interplay between artificial intelligence and information privacy, particularly focusing on the implications of synthetic data (SD) as a form of background knowledge (BK) in compromising privacy. Through the development of a Synthetic Data Vaults tool tailored for social platforms, the project addresses the escalating challenges posed by toxic and adversarial attacks on real data in online communications. By leveraging a novel taxonomy of BK, the project demonstrates the potential risks of SD in facilitating privacy breaches, such as re-identification and sensitive attribute disclosure. Utilizing advanced techniques, including the Conditional Tabular Generative Adversarial Network (CTGAN) algorithm, the project aims to generate resilient synthetic data while mitigating privacy risks. With a front-end built using HTML, CSS, and JavaScript, and a back-end powered by the Node.js framework, the project offers a secure and privacy-preserving environment for storing and managing synthetic data. Furthermore, by integrating real-time data anonymization methods and online learning techniques, the project ensures robust data protection without compromising user privacy. As SD predominantly relies on AI models, the project highlights the potential for amalgamating AI techniques with traditional anonymization methods to enhance defences against SD-based attacks, paving the way for future advancements in adapting to real-world scenarios, integrating with Differential Privacy (DP) models, and facilitating data augmentation for machine learning applications.

REFERENCES

[1] A. Majeed and S. O. Hwang, ”Rectification of Syntactic and Semantic Privacy Mechanisms,” IEEE Security & Privacy, vol. 21, no. 5, pp. 18-32, Sept.-Oct. 2023, doi: 10.1109/msec.2022.3188365.
[2] A. Majeed and S. O. Hwang, ”A Practical Anonymization Approach for Imbalanced Datasets,” IT Professional, vol. 24, no. 1, pp. 63-69, Jan.-Feb. 2022, doi: 10.1109/MITP.2021.3132330.
[3] 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.
[4] 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
[5] 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
[6] A. Majeed and S. O. Hwang, ”When AI Meets Information Privacy: The Adversarial Role of AI in Data Sharing Scenario,” IEEE Access, vol. 11, pp. 76177-76195, 2023, doi: 10.1109/ACCESS.2023.3297646.
[7] L. Breiman, ”Random forests”, Mach. Learn., vol. 45, no. 1, pp. 5-32, 2001. doi: 10.1109/ACCESS.2023.3235969.
[8] Lei Xu, Maria Skoularidou, Alfredo Cuesta-Infante, Kalyan Veeramachaneni, ”Modelling Tabular data using Conditional GAN,” NeurIPS, 2019.
[9] K. Fang, V. Mugunthan, V. Ramkumar and L. Kagal, ”Overcoming challenges of synthetic data generation”, Proc. IEEE Int. Conf. Big Data (Big Data), pp. 262-270, Dec. 2022.
[10] Kim, Taehoon, and Jihoon Yang, ”Selective Feature Anonymization for Privacy-Preserving Image Data Publishing,” Electronics, vol. 9, no. 5, p. 874, 2020, doi: 10.3390/electro.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY